Efficient Texture Descriptors for Localisation of Masses in Local Dense Background of a Mammogram

by

Shelda Sajeev

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy
College of Science and Engineering
July 2019

Contents

7   Automatic Mass Localisation   108

8   Conclusion and Future Work   118

List of Figures

List of Tables

List of Abbreviations

AUC   Area Under the ROC curve

BI-RADS   Breast Imaging Reporting and Data System

BSSA   BreastScreen SA

CAD   Computer-Aided Detection

CLAHE   Contrast Limited Adaptive Histogram Equalization

DDSM   Digital Database for Screening Mammography

FCM   Fuzzy CMeans

FPI   False Positive per Image

FROC   Free-response Receiver Operating Characteristics

LBP   Local Binary Patterns

LDA   Linear Discriminant Analysis

ROC   Receiver Operating Characteristics

ROI   Regions Of Interest

SLBP   Superpixel-based LBP

SRM   Statistical Region Merging

SVM   Support Vector Machine

Abstract

Breast cancer is considered to be one of the major health problems and leading causes of death among women worldwide. Screening mammography is the most used and cost-effective tool for detecting early stage cancer. However, detecting cancerous masses in the dense background of a breast is a particularly challenging task, even for an experienced radiologist. This stems from the similarity of intensity between the masses and the overlapped dense normal tissues. Mammographic sensitivity is less than 50% in women with dense breasts. The need for improved diagnosis of breast cancer in women with dense breast is further emphasized by the greater risk of breast cancer in this population. Women with dense breast have four to five times higher risk of getting breast cancer compared to women with little or no dense tissues. Computer-aided detection (CAD) has been developed to assist radiologists in early breast cancer detection and diagnosis. Although many CAD techniques have been developed for mass classification/detection, the CAD sensitivity in dense breast is still low.

This study aims to improve detection of cancerous masses localised in the dense background of breasts by characterising the textures of masses, based on primitive micropatterns (at pixel level) and their macro level (superpixel) representations. A new paradigm for texture analysis, based on superpixel tessellation, is the main contribution of this thesis. The paradigm enables new mechanisms for understanding complex texture structures in images. Both pixel and superpixel level micropatterns are used in this study to distinguish breast masses from normal dense tissues. The results indicate that the proposed textural features can produce highly effective and efficient descriptors of breast masses, localised in a dense background.

The effectiveness of the proposed approaches is validated on two datasets (DDSM and BSSA) using performance measures such as Dice Index, Hausdorff distance, receiver operating curve (ROC), area under the curve (AUC) and free receiver operating curve (FROC). The experimental results indicate that the proposed methods can classify masses with AUC score up to 0.97 and can localise masses with sensitivity of 80% with only 2.7 false positives per image.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Shelda Sajeev

Refereed Publications from the Study

Referred Conference Paper

1. Shelda Sajeev, Mariusz Bajger, Gobert Lee, 'Superpixel pattern graphs for identifying breast mass ROIs in dense background: a preliminary study', *Proc. SPIE 10718, 14th International Workshop on Breast Imaging (IWBI 2018)*, 107180V, July 2018.

2. Shelda Sajeev, Mariusz Bajger and Gobert Lee, 'Structured Micro-Pattern Based LBP Features for Classification of Masses in Dense Breasts', *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Sydney, NSW, 29 Nov.-1 Dec. 2017, pp. 1-8.

3. Shelda Sajeev, Mariusz Bajger and Gobert Lee, 'Improving breast mass segmentation in local dense background: An entropy based optimization of statistical region merging method', *13th International Workshop on Breast Imaging (IWDM)*, Malmo, Sweden, 19-22 June 2016, pp. 635-642

4. Shelda Sajeev, Mariusz Bajger and Gobert Lee, 'Segmentation of Breast Masses in Local Dense Background Using Adaptive Clip Limit-CLAHE', *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Adelaide, South Australia, 23-25 November 2015, pp. 1-8.

Referred Journal Paper

5. Shelda Sajeev, Mariusz Bajger and Gobert Lee, 'Superpixel Texture Analysis for Classification of Breast Masses in Dense Background', *IET Computer Vision*, 12(6):779-786, March 2018.

Acknowledgements

Chapter 1

Introduction

This introductory chapter provides a brief summary of the incidence of breast cancer, current best clinical practice for early breast cancer detection, and presents the motivation and objectives of this thesis. Section 1.1 describes an overview of breast cancer incidents and Section 1.2 provides an overview of widely accepted clinical practice for early breast cancer detection. The most commonly found mammographic abnormalities are described in Section 1.3. Dense breast tissues and commonly used methods for estimating breast density are described in Section 1.4. Computer-aided detection system is described in Section 1.5. Cancer detection in dense breast is discussed in Section 1.6. Finally, motivation and research objectives of this thesis are summarised in Section 1.7 followed by an overview of the thesis structure in Section 1.8 .

1.1 Breast Cancer Statistics

Breast cancer is considered to be a major health problem and one of the leading causes of death in women for more than five decades (Tang et al. 2009, Sardanelli et al. 2017). The incidence of breast cancer has increased worldwide in recent years. Based on statistics from the World Health Organization (WHO), every year about 14 million people are diagnosed with cancer. By 2025 it is predicted to be 19 million people per year before striking 24 million by 2035. Approximately 1.7 million new breast cancer cases were diagnosed worldwide in 2012, which accounts for a quarter (25%) of all cancers identified in women (WCRFI 2017).

According to the Australian Institute of Health and Welfare (AIHW) and the Australian Cancer Society (ACS), breast cancer is the most commonly diagnosed

cancer among Australian women. It is the second most frequently diagnosed cancer to cause death in women, after lung cancer, accounting for 14.9% of all cancer related mortality in women in 2017. The number of new breast cancer cases is showing an increasing trend according to the cancer survey. In 1982 the number of new cases of breast cancer diagnosed in Australian women was 5,303 and in 2010 it had increased to 14,181. In 2014, it was 15,270 and in 2020 about 17,210 Australian women are expected to be diagnosed with breast cancer. On an average, 1 in 10 Australian women have a chance of developing breast cancer before the age of 75 and 1 in 8 Australian women have a chance of developing breast cancer before the age of 85 (AIHW 2017). Most of the women diagnosed with breast cancer were aged 50 to 69. This shows a greater risk of breast cancer with increase in age. Figure 1.1 shows the details of breast cancer incidence and mortality rates by age of diagnosis and by sex.

Image removed due to copyright restriction.

Figure 1.1: Australian breast cancer incidence and mortality rates, by sex and cancer diagnosed age (AIHWC 2017).

In the United States (US), as per the American Cancer Society, breast cancer is the leading cause of death among women. It was estimated that 266,120 new breast cancer cases were identified, and 40,920 were expected to die from breast cancer among US women in 2018 (ACS 2018). Figure 1.2 shows the details of a cancer diagnosis survey conducted in the United States. The chance of an American woman developing breast cancer in her lifetime is 12.4% (one in eight women).

Also, in Europe, breast cancer is the most common cancer in women, with

Image removed due to copyright restriction.

Figure 1.2: United States cancer incidence rates for most commonly diagnosed cancers for male and female, 1975 to 2010 (Siegel et al. 2014).

1 in 8 women having a chance of developing breast cancer before the age of 85. In 2018, the estimated number of new female breast incidence was 523,000 cases and estimated death was 138,000 (Ferlay et al. 2018).

According to Cancer Research UK, there were 55,122 new cases of invasive breast cancer in 2015, accounting for 15% of all new cancer cases, and 11,433 died from breast cancer in 2014 (CRUK 2015).

1.2   Screening Mammography

Due to this high incidence rate of breast cancer in females compared to other types of cancer, some countries have introduced screening programs for early detection. In Australia, women aged 50-74 years are recommended to participate in the BreastScreen Australia Program every two years for free screening mammograms (BSA 2019). In the United States, women aged 45-54 are recommended

for screening mammography every year and, women aged 55 and older for every two years or they could continue yearly screening (ACSBC 2018). In UK, women aged 50-70 years are invited for mammographic screening every three years (NHS 2018).

Mammography, an x-ray imaging technique, is the best available screening tool for detecting breast cancer at its early stages (Maggio 2004, NHS 2012). It has shown to reduce the mortality from breast cancer by 30-40% (Weedon-Fekjær et al. 2014, ACR 2018). In-addition to screening, mammography can also be used to diagnose cancer (NBCF 2016). The purpose of screening mammography is to determine if there is any indication of early signs of breast cancer. It is done on women who have not reported any signs of breast cancer. Diagnostic mammography is done on women who have found any change or abnormality in the breast during self-examination or have found abnormality during the screening.

As the cause of breast cancer is unknown, the only way to reduce the morbidity and mortality rates associated with breast cancer is early detection through screening mammography. This can give women a greater chance of receiving successful treatment at a very early stage. Figure 1.3 shows the importance of screening. Without screening a tumor is normally found at a stage where it is considerably large in size and may be too late to save the life of the person. This thesis deals with screening mammogram.

Image removed due to copyright restriction.

Figure 1.3: Graph showing the importance of screening (Weedon-Fekjær et al. 2010).

Image removed due to copyright restriction.

Figure 1.4: CC and MLO views (Resource: http://www.imaginis.com).

Image removed due to copyright restriction.

Figure 1.5: Examples of digital and film mammogram (Parker-Pope 2008)

In routine screening mammography, each breast has two mammograms taken to capture information from two different views. One is the Cranio-Caudal (CC) view which is taken from top to bottom and the other is the Medio-Lateral Oblique (MLO) view which is taken from the side. Figure 1.4 shows the CC and MLO mammographic views. The mammograms taken are then interpreted by radiologists. If radiologists find any abnormality, the woman is called back for further investigations. Double reading is the common practice in Australia and Europe (Giordano et al. 2012, Pow et al. 2016, NAS 2008). In double reading, two radiologists will independently read the mammograms. If there are any discrepancies between them, then the mammogram will be further reviewed by a third radiologist or by a consensus panel or committee (Posso et al. 2017).

The mammographic images for examination can be captured in two different ways (Gur 2007, Pisano et al. 2007). One is screen film mammography where the breast images are captured on film cassette. The other is full field digital mammography where digital breast images are captured and stored in a computer. Figure 1.5 shows an example of digital and film mammogram. Although, digi-

tal mammograms have some advantages, like easy access, storage and improved means of transmissions, it has not been proved that the diagnostic accuracy of digital mammography is better than film based. One major disadvantage of digital mammography is the cost, which is 1.5 to 4 times higher than film mammography and, hence, it is not available in most screening centres in developing countries (Pisano et al. 2005).

1.3 Abnormality Indicators in Mammograms

There are different types of mammographic abnormalities. The most commonly detected mammographic abnormalities are microcalcifications and masses. Microcalcifications are tiny deposits of calcium and they appear as small white dots in mammograms. Breast mass is a term used to indicate a localized swelling or lump in the breast (Oliver et al. 2010) and it is defined in mammography as a space-occupying lesion that is noticeable in at least two different mammographic projections (Tang et al. 2009). Masses vary in size and shape like circumscribed, spiculated, lobulated or ill-defined (Oliver et al. 2010), and radiologists examine these associated mass properties to classify them as benign or malignant. It is widely known that mass detection in mammogram is a challenging task. It is particularly challenging when masses are localized in the dense regions of a mammogram. Figure 1.6 shows a mammogram with a mass localized in dense background. It is evident that the mass is hardly visible and the boundaries are extremely hard to detect. This thesis deals with masses localized in dense background.

1.4 Breast Density/Dense Breast Tissue

Breast density is defined as the percentage of the breast occupied by fibroglandular tissue (dense tissue) in relation to the fatty tissue (nondense tissue) (Boyd et al. 1995). The association of breast parenchymal tissues and breast cancer has been studied for three decades (Wolfe 1976a,b, Boyd et al. 2007, Harvey & Bovbjerg 2004, McCormack & Silva 2006, Chen 2013, Tice et al. 2013, Yaghjyan et al. 2013, Xi-Zhao 2014). Studies show that there is an increased risk associated with dense breast tissues. Several authors have concluded that women with dense breast tissues that occupy more than 75% of the mammogram have 4 to 6 fold increased risk of developing breast cancer compared to others who have less or
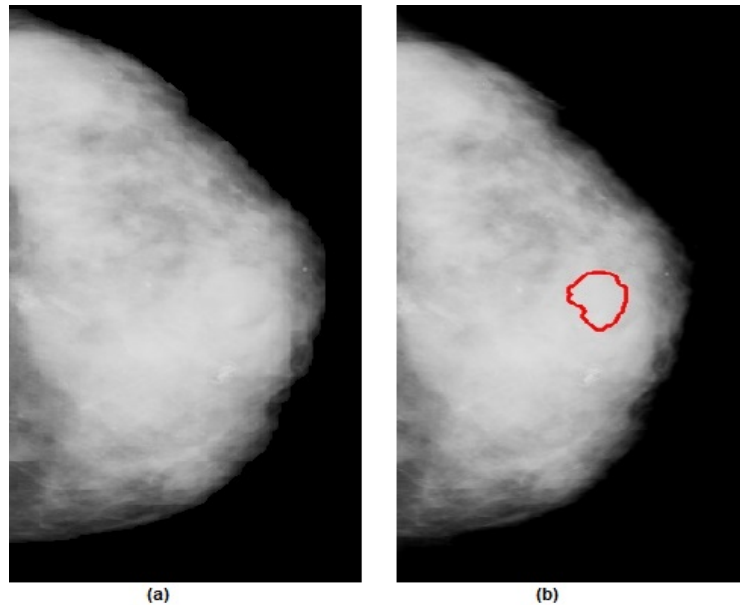
Figure 1.6: Example of mammogram showing mass in dense background (a) without mass outline (b) with mass outline.

nondense tissues (Boyd et al. 1995, 2007, Harvey & Bovbjerg 2004, McCormack & Silva 2006, Ursin et al. 2003, Vacek & Geller 2004, Byrne et al. 1995, Mandelson et al. 2000, Tice et al. 2013, Yaghjyan et al. 2013). The first report for the association of patterns of breast parenchymal tissues and breast cancer was published by Wolfe (1976a). The study described four groups of parenchymal patterns on the basis of radiographic appearance. The four groups are Nl, PI, P2, and DY. The Nl and PI groups were considered to be low risk due to the breast being composed of mostly fatty tissues. The P2 and DY groups were considered to be at high risk for developing breast cancer due to the breast being composed of mostly dense tissues. Modification of Wolfe's parenchymal patterns classification were proposed by Tabar & Dean (1982) and Gram et al. (1997), both giving more focus to structure of the patterns reported in the Wolfe study.

Boyd et al. (1995) and BI-RADS (Breast Imaging Reporting and Data System) classification have given more focus to the percentage of the dense breast tissues instead of the structure of the patterns. Boyd et al. (1995) defined six classifications based on the percentage of dense breast tissues as follows: none, less than 10%, 10 to 25%, 25 to 50%, 50 to 75% and above 75%. BI-RADS classification is introduced by the American College of Radiology (ACR). According to ACR fourth edition (ACR 2003), have four main classes: BI-RADS I (breast is almost fatty, glandularity < 25%), BI-RADS II (scattered fibro glandular densities, glandularity 25-50%), BI-RADS III (heterogeneously dense, glandularity

51-75%)and BI-RADS IV (breast is extremely dense, glandularity >75%). Figure 1.7 shows examples of mammograms from four BI-RADS density classes. The fifth edition (ACR 2014) also have four classification same as previous one, but does not have the percentages.

(a) BI-RADS I

(b) BI-RADS II

(c) BI-RADS III
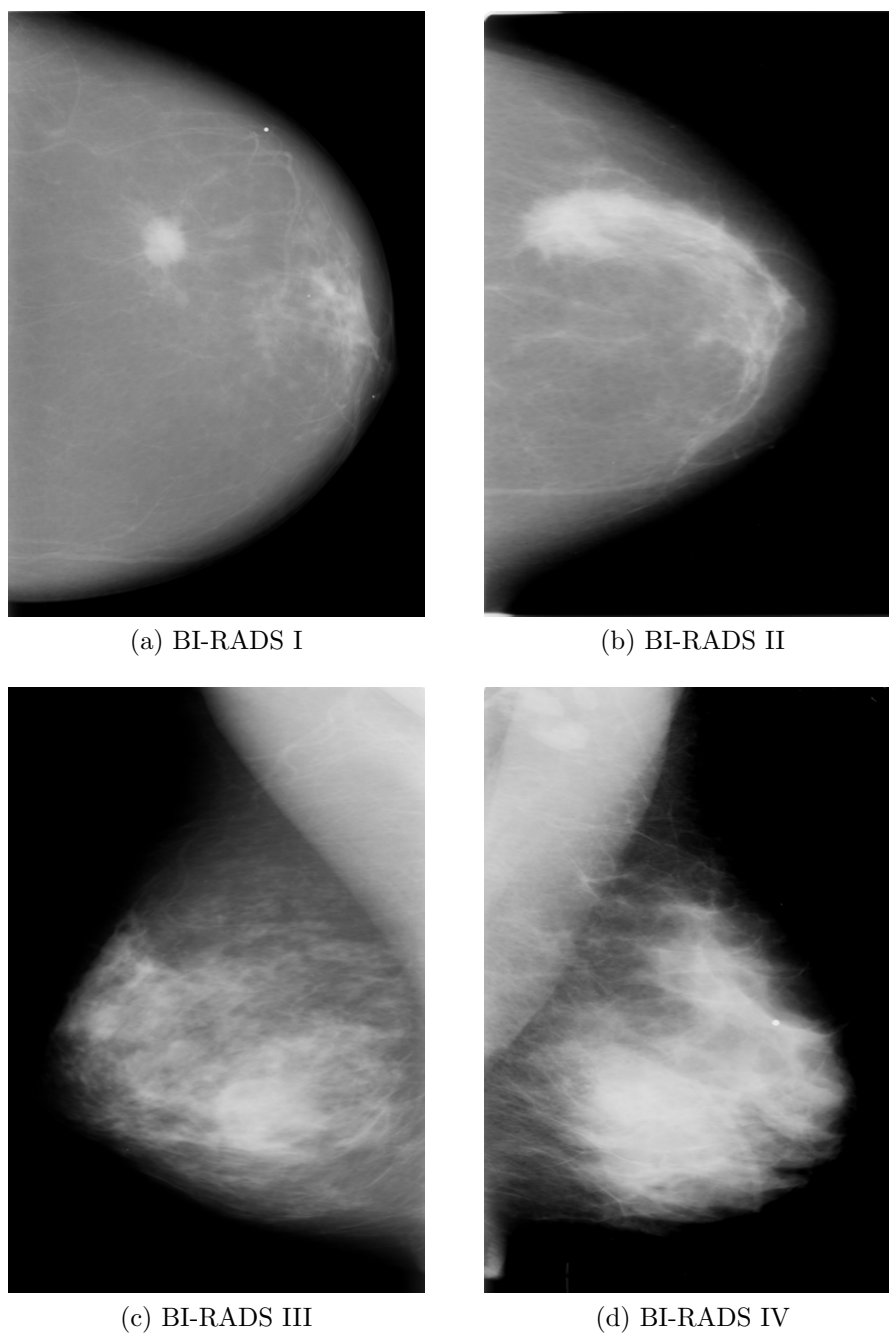
(d) BI-RADS IV

Figure 1.7: Examples of mammograms from four BI-RADS density classes from DDSM database (a) BI-RADS I, (b) BI-RADS II, (c) BI-RADS III and (d) BI-RADS IV.

1.5 Computer-Aided Detection

Computer-Aided detection (CAD) is an ongoing area of research and has attracted a lot of attention over the last two decades. A CAD system integrates digital image analysis techniques, pattern recognition and artificial intelligence techniques to find abnormalities in mammograms (Rangayyan et al. 2007). It was developed to help radiologists in making precise clinical decisions for early detection of breast cancer (Morton et al. 2006). A radiologist still needs to read the mammogram with a CAD system acting as a second reader and making the radiologist's work easier by evaluating the mammogram and highlighting suspicious regions for review (Stark 2000). If two radiologists have different opinions on a mammogram, the CAD system can help them by providing a machine opinion to reconsider.

Many studies have shown that with advances in technology, CAD systems can improve early breast cancer detection. A systematic review (Eadie et al. 2012) of CAD found an increase of 8% and 7% in sensitivity and specificity respectively, when using CAD compared to a radiologist alone. Cupples et al. (2005) reported the results of a large clinical study, conducted with 27,274 screening mammograms. With the help of CAD, there was 8.1% increase in recall rate, 6.7% increase in biopsy rate, and 16.1% increase in cancer detection. Birdwell et al. (2005) conducted a study on 8,682 patients in a university hospital setting and reported 7.4% more cancer detection using CAD. A study by Morton et al. (2006) on 21,349 screening mammograms reported an increase of 10.77% in recall rate and 7.62% in cancer detection rate. In (Brem et al. 2003), 377 screening mammograms were interpreted by three radiologist with and without the help of CAD systems. The study found that CAD significantly improved the detection of breast cancer by increasing radiologist sensitivity by 21.2%. Freer & Ulissey (2001) conducted a study using a database containing 12,860 mammograms to assess the performance of radiologists with and without the assistance of CAD. The performance of the CAD system was measured using recall rate, positive predictive value, cancer detection rate and detection of early-stage malignancies. The study found that with the help of CAD, recall rate increased from 6.5% to 7.7%, an increase of 19.5% in number of cancers detected and early-stage malignancy detection increased from 73% to 78%. There was no change observed in the positive predictive value for biopsy.

Studies have shown that double reading can increase the breast cancer detection rate (Dinnes et al. 2001, Harvey et al. 2003, Helvie 2007) and hence it

is the common practice for interpretation of screening mammograms in many countries (Giordano et al. 2012, Pow et al. 2016, NAS 2008). Many studies suggest single reading with CAD as an alternative (Gilbert et al. 2008, Posso et al. 2017, Sato et al. 2014), since double reading needs two radiologists. In addition, CAD has shown potential to improve the detection rates similar to what double reading can achieve. Gilbert et al. (2008) conducted a trial on 31,057 women undergoing routine screening mammography, to compare the performance of single reading with CAD and double reading. The trial found that single reading with CAD obtained similar cancer detection rate as double reading (87.2% for single reading with CAD and 87.8% for double reading) and a 0.5% higher recall rate compared to double reading (3.9% for single reading with CAD and 3.4% for double reading). Azavedo et al. (2012) was not able to find enough evidence to determine whether the performance of single reading with CAD is equivalent to that obtained in double reading. Sato et al. (2014) shows single reading with CAD as a cost effective approach that can be included in the screening program instead of double reading.

A large variety of CAD approaches for mammographic image analysis can be found in the literature. A review of CAD methods for mammographic mass detection is presented in Section 2.2.

1.6 Cancer Detection in Dense Breast

As mentioned previously, mammography is the best available screening tool for early detection of breast cancer. However, screening effectiveness declines significantly with an increase in overall breast density (Castellano et al. 2011). Mammographic sensitivity may be as low as 30% to 48% in women with dense breasts (Castellano et al. 2011, Kolb et al. 2002).

Kolb et al. (2002) conducted a study on the performance of screening mammography with 27,825 screening sessions. The results show that in women with dense breast, mammography failed to show 52% of cancer while in women with nondense breast, mammography failed to show only 2% of breast cancers. Mandelson et al. (2000) shows that breast density was strongly associated with reduced mammographic sensitivity and increased risk of interval cancer. Interval cancers are those cancers that are detected some time after a mammographic screening in which findings are considered normal. The mammographic sensitivity declined sharply from 80% among women with predominantly fatty breasts to 30% in

women with extremely dense breasts in their experiment. A study published by Buist et al. (2004) indicates that breast density is the major factor contributing to the cancers being undetected at screening.

The need for improved diagnosis of breast cancer in women with dense breast is further emphasized by the greater risk of breast cancer in this population. According to Boyd et al. (2002) women with dense breast have 1.8 to six fold increased risk of developing breast cancer compared to those who have less or no dense breast tissues. Mandelson et al. (2000) suggest that women with dense breast have six fold greater risk of interval cancer. Hou et al. (2013) also reported that women with dense breast have high risk of breast cancer.

CAD systems, which have been developed to help radiologists in making precise clinical decisions for early detection of breast cancer (Morton et al. 2006), may also often fail to perform in dense breasts (Ho & Lam 2003, Brem et al. 2005, Obenauer et al. 2006, Malich et al. 2006, Oliver et al. 2010, Castellano et al. 2011, Liu et al. 2011, Manso et al. 2013, Tai et al. 2014, de Oliveira Silva et al. 2017). Christoyianni et al. (2000) presented a detection algorithm for circumscribed masses in mammograms using radial basis function neural network (RBFNN). The experiment was tested on 22 mammograms containing circumscribed lesions from Mini-MIAS. The authors showed that the recognition of abnormal tissue was 90.9% in fatty tissue while it dropped to 33.3% in dense tissue. Ho & Lam (2003) showed a decrease in sensitivity of CAD from 93.9 % in women with nondense breasts to 64.3 % in women with dense breasts. Liu et al. (2011) study based on Complete Local Binary Pattern (CLBP) achieved a sensitivity of 81% with 1.78 false positives per image for BI-RADS I but dropped to a sensitivity of 33.3% with 1.33 false positives per image for BI-RADS IV. The method was evaluated on a dataset with 231 images, containing 245 masses taken from DDSM database. Tai et al. (2014) study based on co-occurrence matrix and optical transformation methods achieved a sensitivity of 90% with 2 false positives per image for BI-RADS I, while it dropped to 50% with 2 false positives per image for BI-RADS IV. The experiment was conducted on 358 mammograms selected from DDSM. Manso et al. (2013) analysed the effect of breast density on detection of masses. In the study, mass was detected using an Independent Component Analysis (ICA) feature extraction method and classification by neural networks (NN) and SVM classifiers. The experiment was tested on 2324 mammograms taken from DDSM. It was found that breast density affects the performance, since the performance measure area under curve decreases from 0.965 to 0.892 (-7.56%) for NN classifiers and 0.964 to 0.897 (-6.95%) for SVM classifiers, when moving from BI-RADS I

to BI-RADS IV. Recently, de Oliveira Silva et al. (2017) also showed that masses in low density are better detected (accuracy of 92.71% in nondense breasts) than masses in high density (accuracy of 79.17% in dense breasts).

Few studies (Pandey et al. 2012, Singh & Bovis 2005) suggested image enhancement can increase the contrast between the malignant and normal tissues and there by facilitate mass detection in dense breast. Vallez et al. (2014) and Freixenet et al. (2008) studies suggested that breast tissue classification (or having breast density information) prior to mass detection could help to tune the experiments based on different densities and there by increase mass detection in dense breast.

1.7 Motivation and Research Objectives of the Thesis

It is clear from the literature that dense breast is an important risk factor for breast cancer and mammography is limited in detecting cancer in the dense breast. Even though there are many CAD techniques developed for mass classification/detection, the sensitivity in dense breast is still low. Thus, further research is required to improve the sensitivity of mammography for detection of masses in dense breasts. Therefore, the objective of this thesis is to develop computer methods for analysing masses localized in dense background of mammograms in order to improve the mass detection in dense breast.

Relative location of mass to the dense regions of the breast has a major influence in detectability of the mass. Masses localized in dense background are difficult to identify, even for experienced radiologists as normal dense breast tissues hides the tumour (masses). Different types of breast tissues respond differently to X-rays. Fatty tissues are radiographically translucent and let most X-rays pass through during the mammographic imaging process and, therefore, appear dark in mammographic images. In contrast fibroglandular tissues are radiographically dense and absorb more X-rays than fatty tissue during the acquisition process and, therefore, appear brighter in mammographic images (Wolfe 1976b). The breast tumors are also dense tissues and they appear as solid white areas on the mammograms which makes it difficult to detect tumors in dense breasts. Figure 1.8 shows examples of mammograms with masses localized in dense and non dense backgrounds. It is clear from the figure that detection of masses localized

in dense background is a challenging task for human eyes compared to mass detection in non dense background. This study specifically focused on analysing masses localized in the dense background of mammograms.
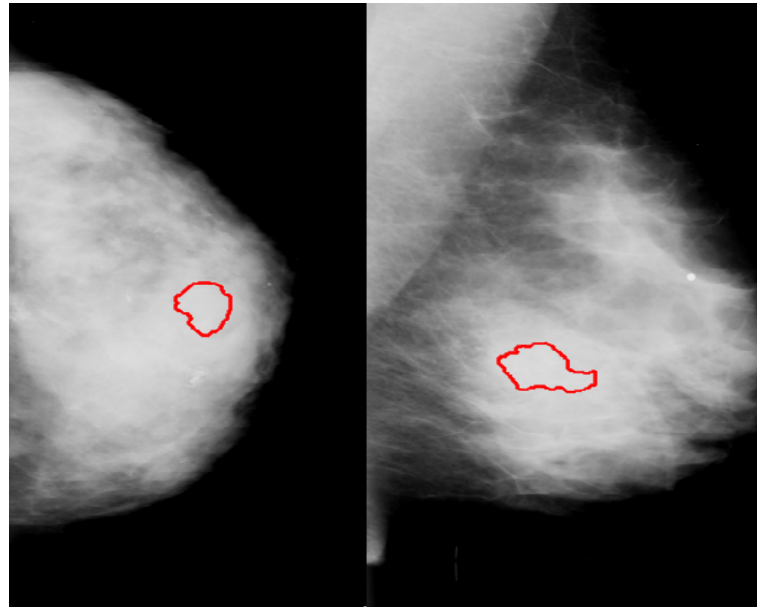
(a) Masses in dense background
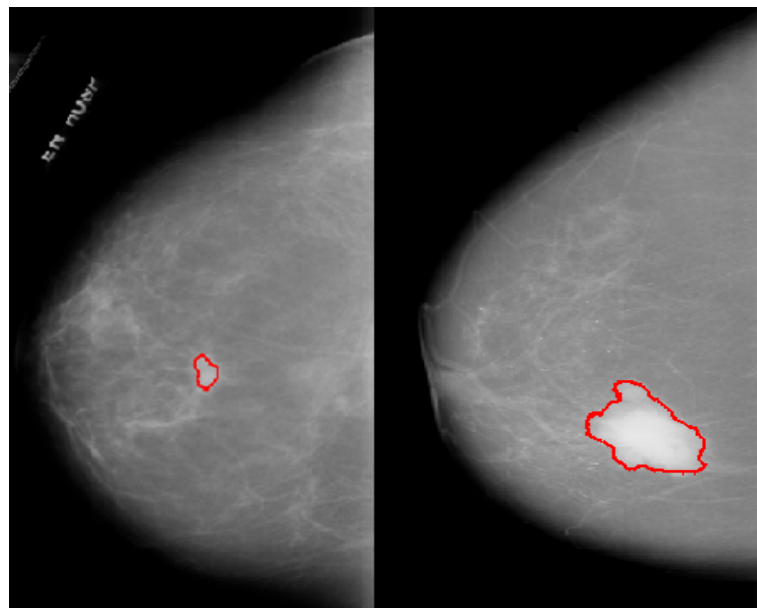
(b) Masses in non-dense background

Figure 1.8: Examples of mammograms (a) showing masses in dense background and (b) masses in non dense background.

The main objective of this thesis is to develop texture analysis methods suitable for improving detection of masses in dense regions of a mammogram. Texture analysis of primitive micropatterns (at pixel level) and their macro level (super-

pixel) representations and suitable mechanisms to understand their topological representation in masses will be studied. In addition, methods to improve breast mass segmentation in dense breast, which is an important step for subsequent mass detection process, will be investigated. Finally, an automatic breast mass localisation system incorporating the methods is developed.

1.8   Overview of the Thesis

This thesis consists of 8 chapters including this introduction chapter (Chapter 1).

Chapter 2 provides a review of the main steps involved in an automated CAD system and describes key methods used in this thesis for developing the final automatic mass localisation system. It also provides an overview of the currently available texture analysis techniques and their applications to mass classification/detection.

In Chapter 3, characteristics of masses localized in dense background are studied using primitive/structured micropatterns generated using Local Binary Pattern (LBP) technique. The typical histogram concatenation approach of extracting features from LBP is replaced by features based on combinatorial properties of the structured micropatterns. This allows for a huge reduction of the dimension of the feature space while keeping a high rate of mass classification.

Chapter 4 presents a new paradigm for texture analysis: the superpixel texture analysis paradigm. In this paradigm, the ROIs are transformed into a superpixel tessellation using a regular grid approach and irregular grid approach. Superpixel patterns are found using a generalized superpixel based local binary pattern technique and density of macro-structures in the ROI image form the basis for classification.

Chapter 5 presents a method for building graphs on the structured superpixel patterns. The topology/connectivity of the structured superpixel patterns is analyzed using multiscale morphology. Graph models are constructed on structured superpixel patterns using morphological dilation to represent the spatial connectivity relationship between the structured superpixel patterns within an ROI. A set of graph features is generated from the proposed model and classification is performed based on graph features.

In Chapter 6, two techniques for improving breast mass segmentation in local dense background are presented. Initially, a self-adjusting mammogram contrast

enhancement solution called adaptive CLAHE is proposed, aiming to improve mass segmentation. Then, an optimization algorithm for automatically tuning segmentation of mammograms by the Statistical Region Merging (SRM) technique is proposed, to improve the mass segmentation. Results are compared with commonly used image enhancement techniques in the literature.

Chapter 7 presents an automatic mass localisation system incorporating the methods developed in the previous chapters (Chapters 3, 4, 5 and 6).

In Chapter 8, conclusions, future work and contributions of this thesis are summarized.

Chapter 2

Technical Background and Literature Review

In this chapter we review the main steps involved in an automated CAD system and provide an overview of the currently available texture analysis techniques and their applications to mass detection/classification. The databases searched for the literature were ScienceDirect and Google scholar. Reference lists of the interested papers were hand searched for additional records. In addition, manual searching from proceedings of top conferences in Breast Imaging like IWDM, MICCAI and citations of key authors. The keywords used for literature search were mammography, mass detection/classification, CAD, texture analysis, superpixel, graph model, micro-patterns, local binary patterns, dense breast, mass enhancement, and mass segmentation. Texture analysis methods and their applications to mass detection/classification are presented in Section 2.1. A brief review of the main steps in the mammographic mass detection and the key methods that have been incorporated into the mammographic CAD mass detection system developed in this thesis are described in Section 2.2. An overview of the databases used in this thesis is described in Section 2.3.

2.1   Texture Analysis

Texture is an important feature of images, formed by variations in the grey levels of neighbouring pixels of the image. It can be defined as a repeated pattern of fundamental units called Texels (Blostein & Ahuja 1989, Hild & Shirai 1993). Texture features have been used and proven to be useful in discriminating different classes in many applications of computer image analysis. In mammographic

image analysis, texture features have been widely used for characterizing breast tissue patterns (Rabidas et al. 2018a, Suhail et al. 2017, Muramatsu et al. 2016, Raghavendra et al. 2016, Abdel-Nasser et al. 2015, Wei et al. 2012, He et al. 2011, Blot & Zwiggelaar 2005). Tables 2.1 and 2.2 provide a summary of texture methods for mass classification and detection respectively, indicating the texture features used and their results. The methods for extracting texture features can be divided into two main categories: statistical and structural. The following subsections briefly explain statistical and structural methods.

2.1.1 Statistical Methods

Statistical methods characterize textures based on the spatial distribution of intensity values in an image. Statistical methods can be classified into three main categories: first order, second order and higher order statistics. The first order statistics estimates the texture properties based on individual pixel values, disregarding the relationship between adjacent pixels. Average, variance, skewness and kurtosis computed on an image intensity histogram are some of the first order measures. The second order and higher order statistics estimate the texture properties based on the relationship between two or more pixels. Gray level co-occurrence matrices (GLCM) proposed by Haralick et al. (1973) is a popular method for extracting second order texture information from the images. It is also referred to as spatial gray level-dependence (SGLD) matrices (Rangayyan 2005). Haralick described fourteen statistical measures (like angular second moment, contrast, entropy, correlation, and inverse difference) from GLCM to represent the characteristics of the spatial distribution of gray levels in the image. Gray level run length method, a higher order statistic, proposed by Galloway (1975), is another popular statistical texture analysis method. A gray level run is a set of consecutive pixels having the same intensity/gray value and run length is the number of pixels in a run. Short and long runs emphasis, gray level and run length non-uniformity and run percentage are some of the texture features extracted from gray level run length matrix. These methods are widely used in analysis and classification of breast masses (Chan et al. 1995, Wei et al. 1997, Sahiner et al. 1998, Rangayyan et al. 1997, Sahiner et al. 2001, Mudigonda et al. 2001, Khuzi et al. 2009, Amroabadi et al. 2011, Dhungel et al. 2017, Khuzi et al. 2009, Dhahbi et al. 2018, Chakraborty et al. 2018).

Sahiner et al. (1998) introduced a rubber band straightening transform (RBST) method to analyze the margin characteristics of masses. This method maps the

Table 2.1: A summary of texture features for mass classification, indicating authors, year of publication, features used, experimental data, results.

Mass Classification			
Author & Year	Features/Technique	Dataset	Results
Sahiner et al. (1998)	SGLD and RLS features computed using RBST method	249 mammograms	benign/malignant classification AUC = 0.94
Khuzi et al. (2009)	GLCM features	100 mammograms from Mini-MIAS	malignant/normal classification AUC = 0.84
Llado et al. (2009)	LBP features	1792 manually cropped ROIs	malignant/normal classification AUC = 0.94
Choi & Ro (2012)	LBP features	automatically generated 1693 ROIs	malignant/normal classification AUC = 0.92
Torrents-Barrena et al. (2014)	Gabor wavelet filters	322 mammograms from Mini-MIAS	Accuracy = 0.80
Hussain et al. (2014)	Gabor filters	1024 manually cropped ROIs	malignant/normal classification AUC = 0.99
Abdel-Nasser et al. (2015)	Uniform local directional patterns	312 manually cropped ROIs	AUC= 0.93
Muramatsu et al. (2016)	Radial local ternary patterns	376 manually cropped ROIs	benign/malignant classification AUC = 0.90
da Rocha et al. (2016)	LBP and diversity index	1155 ROIs (625 malignant and 530 benign)	AUC =0.88
Raghavendra et al. (2016)	Gabor filter	690 mammograms from DDSM	benign/malignant classification accuracy = 98%
Suhail et al. (2017)	Texton based approach	400 mammograms from DDSM	benign/malignant classification accuracy = 96%
Rabidas et al. (2018a)	Neighborhood structural similarity method	58 mass ROIs from Mini-MIAS and 1316 mass ROIs from DDSM	benign/malignant classification AUC = 0.98 for Mini-MIAS, AUC = 0.93 for DDSM
Dhahbi et al. (2018)	GLCM, fractal dimension, wavelet, Hilbert's image representation and Kolmogorov Smirnov distance	1914 manually cropped masses, automatically generated 8254 normal tissues from DDSM	Accuracy = 81.09
Pawar et al. (2018)	LBP based on sparse curvelet subband coefficients	automatically generated 381 suspicious ROIs from Mini-MIAS and 1343 from DDSM	AUC = 0.98 for Mini-MIAS, AUC = 0.98 for DDSM

group of pixels surrounding a mass into a rectangular array. Their results show that SGLD and run-length statistics (RLS) features computed from RBST transformed images have higher potential for discrimination of the malignant tissues from benign in mammograms, when compared to the same features computed from entire an ROI containing the mass. Mudigonda et al. (2001) used an adaptive ribbons technique to extract pixels surrounding the margin of masses and computed Haralick features to classify malignant and benign regions in mammograms. Chakraborty et al. (2013, 2018) introduced two angle co-occurrence

Table 2.2: A summary of texture features for mass detection, indicating authors, year of publication, features used, experimental data, results.

Mass Detection			
Author & Year	Features/Technique	Dataset	Results
Mudigonda et al. (2001)	Haralick features	56 mammograms from Mini-MIAS	Sensitivity of 81% with 2.2 false positive per image
Liu et al. (2011)	Complete Local Binary Pattern	231 mammograms	Sensitivity = 81% for BI-RADS 1, Sensitivity = 33.3% for BI-RADS IV
Dheeba et al. (2014)	Laws features	216 mammograms from Mini-MIAS	AUC = 0.96, sensitivity = 94% and specificity = 92.1%.
Tai et al. (2014)	Co-occurrence matrix and optical density transformation method	358 mammograms from DDSM	Sensitivity = 90% with 2 false positive per image for BI-RADS I, sensitivity = 50% with 2 false positives per image for BI-RADS IV
Dhungel et al. (2017)	GLCM	410 mammograms from INbreast	Sensitivity = 0.98, specificity = 0.7
Chakraborty et al. (2018)	GLCM and ACM	450 benign, 440 malignant, 410 normal images from DDSM	sensitivity = 85% with 1.2 FPI for detection and AUC = 0.87 for classification

matrices (ACMs) based on the magnitude and angle responses of the Sobel operation and extracted Haralick features from it. Their results show that Haralick features computed from the ACMs have higher potential for mass classification, when compared to the same features computed from GLCM. Recently, Dhungel et al. (2017) used texture features generated from GLCM combined morphological features to pre-train the deep learning Convolutional Neural Networks (CNN) for mass classification. Kanadam & Chereddy (2016) used a sparse matrix to represent the irregular shape of mass and characterize using GLCM and gray level aura matrix (GLAM), in contrast to the traditional fixed size window approach.

Texture analysis based on filters is another popular technique that has been widely used in mammogram image analysis (Wei et al. 1997, Buciu & Gacsadi 2011, Zheng 2010, Hussain et al. 2014). Laws texture energy measures, proposed by Laws (1980), uses 5 filters (level, edges, spots, ripples and waves) to extract micro-structure characteristics of the image, which can be used for texture classification. Miller & Astley (1992) used Laws texture energy measures for classifying breast tissues. The study found that edge and spot filters have better texture classification capability than others. Dheeba et al. (2014) used Laws texture features along with an optimized wavelet neural network based on particle swarm optimization to increase the mass classification accuracy.

Other popular filters used for texture analysis are Fourier domain filters, Ga-

bor filters and wavelength transform. From the filtered images, a variety of texture features like edge frequency, randomness, coarseness, linearity, and directionality can be extracted. Fourier analysis requires images to be first transformed into the Fourier domain and then filters are applied on the transformed images. The Fourier transform method extracts global frequency information from signal, however, it takes no account of spatial localisation. Hence, this approach results in poor performance in many applications (Mirmehdi 2008). The general solution is to include spatial dependency into the Fourier analysis via a window Fourier transform. When the Gabor function is used as the window function, it becomes the popular Gabor transform. Gabor filters and wavelet transforms are popular for multi-resolution/multi-scale texture analysis. Zheng (2010) introduced a Gabor Cancer Detection (GCD) algorithm for breast cancer detection. Their Gabor filter was formed with five bands and four orientations (horizontal, vertical, 45 and 135 degree). For each mammogram, 20 filtered images were produced and edge histogram features were generated from these filtered images. Then these features were used for classifying mass and non-mass regions in mammograms. Buciu & Gacsadi (2011) used Gabor features for classifying breast tissues into normal, benign and malignant. The study used 24 Gabor filters which resulted in a 86400-dimensional feature vector for each image. Torrents-Barrena et al. (2014) used multichannel Gabor wavelet filters to extract texture features for classifying tumor and non tumor. Hussain et al. (2014) used Gabor filters along with SVM for false positive reduction and mass classification. They applied a Gabor filter bank at different scales and orientations and extracted statistical features (mean, standard deviation and skewness) from the Gabor images. Raghavendra et al. (2016) have used directional textural features generated using a Gabor filter to differentiate between normal, benign and malignant ROIs.

Wei et al. (1997) used Daubechies' wavelet transform to decompose mammogram ROIs into different scales and computed SGLD features from each scale to form the feature vector for mass classification. Midya & Chakraborty (2015) used Haar wavelet transform to decompose the mammogram region into three levels and computed angle co-occurrence matrices (ACMs) and gray level co-occurrence matrices (GLCM) for each level. Texture features extracted from ACMs and GLCM were used for mass classification.

Fractal theory proposed by Mandelbrot (1977) is another texture analysis method that has been found useful in mammograms. Ke et al. (2010) used the differential box-counting method to derive the fractal dimensions of mammograms. Mammograms were analyzed at multiple scales and fractal features were

extracted from each scale and combined to form a feature vector for mass detection. Zhen & Chan (2001) used a mass detection framework that combines fractal dimension for selecting the ROIs followed by segmentation using multi-resolution Markov random field and classification using shape features.

Local Binary Patterns (LBP) is one of the most discriminative and computationally simple local texture descriptors and has been investigated by researches for about two decades (Ojala et al. 1996, Pietikainen & Zhao 2015). LBP and variants of LBP have been widely used in mammographic mass classification. Llado et al. (2009) and Choi & Ro (2012) used LBP to represent textural properties of masses for reduction of false positives in computer-aided detection of breast masses on mammograms. Muramatsu et al. (2014, 2016) used radial local ternary patterns (RLTP) to differentiate between benign and malignant ROIs. Abdel-Nasser et al. (2015) used uniform local directional patterns (ULDP) to classify breast tissues. Rabidas et al. (2018a) proposed a neighborhood structural similarity (NSS) method based on gray level difference between the adjacent regions of masses. The features generated were combined with uniform LBP features as the NSS features are global in nature. Pawar et al. (2018) proposed an LBP method based on sparse curvelet coefficients to describe masses to reduce the false positives in CAD. Even though LBP based approaches are simple and efficient texture operators, they produce a very large dimension feature space as the features are based on concatenation of histograms of LBP labels.

In this thesis, we propose a novel approach for dense breast mass classification utilizing specific structures of LBP generated patterns, the structured micro and macro patterns. The typical histogram concatenation approach of extracting features from LBP is replaced in our study by features based on combinatorial properties of the structured micro and macro patterns. This allows for a huge reduction of the dimension of the feature space while keeping a high rate of mass classification.

2.1.2 Structural Methods

Structural/geometrical methods are based on the structural properties of texture primitives and understanding the placement rules of these primitives, which governs texture spatial organization (Haralick 1979). Mathematical morphology, edge detection, spot detection and neighbourhood operators are some methods by which primitives can be extracted. Computing statistics of the primitives (e.g. area, intensity, orientation and elongation) and interpreting the placement rules

of the elements are common analysis approaches. Ahuja (1982) and Tuceryan & Jain (1993) used Voronoi tessellations to represent structure and organization of primitives. Zucker (1976) used a graph method to describe spatial relationship between the texture primitives. Lu & Fu (1978) used tree grammar for discrimination of textures.

A large variety of graph based structural approaches can be seen in histopathological image analysis where the spatial connectivity relationship between the cells are measured to distinguish between the healthy and cancer tissues (Gunduz et al. 2004, Oztan et al. 2013, Gunduz et al. 2004, Bilgin et al. 2010, 2007). These graph approaches combine techniques from graph theory, image analysis and machine learning to predict the underlying functional state. The cell graph approaches are based on the assumption that cells in a tissue are organized in a way to perform a specific function and if abnormality is present, that can change this organization. Hence, understanding the structural organization of cells present in the tissues can help to predict the normality or abnormality (Yener 2016). Likewise, in mammogram image analysis, understanding the spatial organization of the structures present in masses can help in predicting cancerous and normal tissues. In a recent study, Cheng et al. (2010) used spatial connectivity graphs to understand the spatial organization of microcalcifications. Graph features extracted from the spatial connectivity graph models were used for the classification of malignant and benign microcalcification clusters. However, the literature lacks structural approaches for mammographic mass texture analysis. The structural approach better suits describing macro-textures, finding a suitable mechanism to extract the macro-structures and their spatial organization information (Zhang & Tan 2002) may be the reason for not using this approach in mass analysis.

In this thesis, we have introduced a novel structural approach for mass analysis where ROIs were first transformed into superpixel tessellation and macro-textures were extracted using a superpixel based local binary pattern technique and the spatial organization of the macro-textures were studied using a graph approach.

2.2   Computer-Aided Breast Cancer Detection

This section provides a brief review of the main techniques that are used for CAD and reviews the key methods used in this thesis. A typical CAD system consists of stages like preprocessing, segmentation, feature extraction, feature selection and classification.

2.2.1 Mammogram Preprocessing

Mammographic images do not provide good contrast between normal dense tissues and cancerous tissues, as X-ray attenuation between these two tissues does not vary much. Studies have shown that if region brightness is less than 2% from its surroundings, then the region is not distinguishable by human eyes (Morrow et al. 1992). This fact is seen quite evidently in the case of masses localized in the dense background of mammograms and this makes it more difficult for the radiologist to distinguish between normal and cancerous tissues. Hence, it is very important to enhance the contrast between the region of interest and background, in order to extract mass features effectively and detect/classify the masses more accurately.

The main task of the mammogram preprocessing stage is to enhance the contrast between the diagnostic features from the background (like contrast between masses and normal breast tissues) using different image processing techniques. Some of the commonly used contrast enhancement techniques are: wavelength transform, unsharp masking, histogram equalization (HE) and contrast limited adaptive histogram equalization (CLAHE) (Cheng et al. 2006). In the wavelet transformation approach, the mammograms are first transformed into a wavelet space by applying a mother wavelet function. Then the wavelet coefficients are modified to enhance the mammographic mass features and finally the enhanced image is reconstructed using an inverse wavelet transformation (Cheng et al. 2006, Mencattini et al. 2008). Unsharp masking is a method used to sharpen the mammogram images by emphasizing the high-frequency portions of the image that contain fine details. It is obtained by subtracting a low-pass filtered image from the input image, which corresponds to a high-pass filtered image (Panetta et al. 2011, Cheng et al. 2003, McSweeney et al. 1983).

HE is one of the simplest method used to enhance the contrast of an image based on adjusting image intensity values. The objective of this method is to transform the image histogram to a uniform one, where the output image has all its gray levels with equal probability of incidences (Rangayyan 2005). Let I be a mammogram image with size P and gray levels $i \in \{0, 1, \cdots L-1\}$. The probability of occurrence of gray level i in the image is $p_i = n_i/P$, where n_i is the number of occurrence of gray level i in I. Then the original image intensity $I(x, y)$ can be mapped into new histogram equalized $G(x, y)$ as follows (Cheng

et al. 2006):

$$G(x, y) = (L - 1) \times \sum_{i=0}^{I(x,y)} p_i.$$ (2.1)

The histogram equalization is a global approach that is applied to the entire image and hence it may not enhance local details of the mammograms. Mostly, in mammogram analysis, it is necessary to enhance the local details of an image. In such situations, local operations are preferred over global operations. The local operation requires the information of local neighborhood, which is specified by a window. Adaptive histogram equalization (AHE) is a variant of HE which works over a local region. However, in AHE, there is a possibility of image over-enhancement as noise is also enhanced.

CLAHE proposed by Pizer et al. (1987) is a modification of AHE algorithm. The algorithm subdivides the image into equal sized nonoverlapping regions and applies HE to them. In contrast to AHE algorithm, CLAHE reduces the noise amplification by clipping the histogram at a user defined value called clip limit. The clipped pixels are redistributed equally among all histograms bins. CLAHE has been used as a preprocessing step for mass segmentation, classification and detection. Studies have shown that CLAHE can improve the segmentation accuracy of mass detection in dense breasts compared to the segmentation obtained from an unenhanced original image (Singh & Bovis 2005). In this thesis, a modified version of CLAHE (adaptive CLAHE) is proposed and used as a preprocessing step to improve mass candidate segmentation in Chapter 6 and Chapter 7. In the proposed approach, instead of user selecting the CLAHE parameters, an optimization algorithm based on entropy is used to automatically select the CLAHE parameters, then standard Fuzzy C-means clustering is used for segmentation.

2.2.2 Mammographic Image Segmentation

Image segmentation is one of the fundamental steps that helps in analysing the image data (Sonka et al. 2007). It is the process of dividing an image into several regions or objects based on some specific similarity measures. The similarity can be in terms of brightness/intensity, colour, texture etc (Gonzalez & Woods 1992, Sonka et al. 2007). For CAD, image segmentation can help in delineation and detection of regions of interest. It can help in extracting the breast region from the background and can also help in detecting mass regions based on the segmented bright regions in mammograms. A large variety of image segmentation algorithms

can be found in the literature. Thresholding, region based segmentation and clustering are some of the commonly used segmentation methods.

2.2.2.1 Thresholding

Thresholding is a simple and popular method used for image segmentation that divides the image into several regions based on image intensity values. It can be divided into global and local thresholding. Global thresholding uses a single threshold for the whole image (Brzakovic et al. 1990) while local thresholding uses multiple thresholds depending on the local characteristics of the image (Kallergi et al. 1992). Selection of the correct threshold value is very important for threshold based image segmentation. One of the simplest threshold selection techniques is detection of image histogram peaks to find the segmentation threshold. Some other automatic threshold selection methods are based on optimizing a criterion function like minimizing the total misclassification error (Nakagawa & Rosenfeld 1979) or maximizing posterior entropy of the partitioned image (Kapur et al. 1985).

Otsu (1979) proposed an automatic threshold selection technique based on minimizing the weighted intra-class variance. The Otsu algorithm has been widely used in medical image segmentation applications. In this thesis, Otsu thresholding is used to separate the breast region from the background (Chapter 7). The algorithm is described as follows. Suppose I is a mammogram image with size P and gray levels $i \in \{0, 1, \cdots L - 1\}$. Let S_0 and S_1 are two classes representing background and foreground, respectively separated by a threshold t, then the class probabilities are defined as follows:

$$w_0(t) = \sum_{i=0}^{t} p_i \quad \text{and} \quad w_1(t) = \sum_{i=t+1}^{L-1} p_i, \tag{2.2}$$

where $p_i = n_i/P$ is the probability of occurrence of gray level i in the image and n_i is the number of occurrences of gray level i in I. The corresponding class means are estimated as $\mu_0(t) = \sum_{i=0}^{t} ip_i/w_0(t)$ and $\mu_1(t) = \sum_{i=t+1}^{L-1} ip_i/w_1(t)$. Similarly, the individual class variances σ_0^2 and σ_1^2 are defined as follows:

$$\sigma_0^2(t) = \sum_{i=0}^{t} [i - \mu_0(t)]^2 p_i/w_0(t) \quad \text{and} \quad \sigma_1^2(t) = \sum_{i=t+1}^{L-1} [i - \mu_1(t)]^2 p_i/w_1(t). \tag{2.3}$$

Finally, the weighted intra-class variance is defined as $\sigma_w^2(t) = w_0(t)\sigma_0^2(t) +$

$w_1(t)\sigma_1^2(t)$. The optimal threshold is obtained by searching through all the t values and selecting the value that minimizes the $\sigma_w^2(t)$. A detailed survey of thresholding techniques and threshold selection methods can be found in (Sahoo et al. 1988, Weszka 1978).

2.2.2.2   Region based segmentation

Region based segmentation divides an image into regions based on some homogeneity property (Gonzalez & Woods 1992, Oliver et al. 2010, Sonka et al. 2007). The homogeneity can be expressed in terms of intensity values, texture or color. Region growing and region split and merge are two basic region based segmentation methods. The region growing method starts with a set of seed points and grows by merging neighbouring pixels that have properties similar with the seed point. The merging procedure stops when there is no pixel to merge that satisfies the homogeneity criterion. The split and merge method starts with splitting the image into sub-regions until all regions satisfy a homogeneity criterion and then adjacent regions are merged if they satisfy another homogeneity criterion.

Statistical Region Merging (SRM) technique proposed by Nock & Nielsen (2004) is a region growing algorithm. In this thesis, SRM is used to generate mass candidates in Chapter 6 and superpixels in Chapters 4, 5 and 7. The SRM algorithm is based on probability theory and contains two components: a merging predicate that decides whether two regions R_1 and R_2 should be merged or not, and the order followed in testing this predicate for growing regions. The merging predicate is defined as follows:

$$P(R_1, R_2) = \begin{cases} true, & \text{if} \quad |\bar{R}_1 - \bar{R}_2| \leq \sqrt{b^2(R_1) + b^2(R_2)} \\ false, & \text{otherwise}, \end{cases} \tag{2.4}$$

where

$$b(R_i) = g\sqrt{\frac{1}{2Q|R_1|}\ln\frac{2}{\delta}}. \tag{2.5}$$

where $\bar{R}_1$ and $\bar{R}_2$ denote the average intensity of the regions R_1 and R_2 respectively, g is the number of the gray levels of the image. Q is a parameter, which is related to the coarseness of the segmentation. δ denotes the probability error and $|.|$ stands for cardinality. The order of testing the predicate is based on the

function f which is defined as follows:

$$f(x, x') = abs(x_{int} - x'_{int}).$$

(2.6)

where x and x' are the pixels of the image I, and $I(.)$ is the intensity function.

2.2.3 Clustering

Clustering associates the data points into groups called clusters based on their similarity. Some of the commonly used clustering methods are K-means, Fuzzy C-means (FCM) and Simple Linear Iterative Clustering (SLIC). K-means clustering (MacQueen 1967), which is the simplest among clustering methods, partitions the given image into k clusters. Given a set of data points, K-means clustering partitions the data points into k clusters by minimizing the objective function O_{kmeans}.

$$O_{kmeans} = \sum_{i=1}^{k} \sum_{j=1}^{S_i} ||x_j - c_i||^2,$$

(2.7)

where x represents the data points, $S = \{S_1, S_2 \cdots S_k\}$ represents clusters and c_i is the mean of points in S_j.

FCM clustering is a modification of K-means clustering, which allows a data point to associate with two or more clusters with varying degrees of membership. In this thesis, FCM is used to generate mass candidates in Chapters 6 and 7. Let $X = \{x_1, x_2 \cdots x_n\}$ represents an image with n pixels. The algorithm tries to partition the image into k clusters by minimizing the objective function O_{FCM}.

$$O_{FCM} = \sum_{i=1}^{n} \sum_{j=1}^{k} U_{ij}^m ||x_i - c_j||^2,$$

(2.8)

where m is the parameter controlling the fuzziness, U_{ij} is the degree of membership of data point x_i in the cluster j, c_j is the value of jth cluster center, $||.||$ is the similarity measure between data point and cluster center.

SLIC was proposed by Achanta et al. (2012). In this thesis, SLIC is used to generate superpixels in Chapter 4. SLIC begins by sampling the cluster centers on a regular grid spaced S pixels apart. It is a modified form of K-means clustering where the size of the search window is limited to $2S \times 2S$ around the cluster center to speed up the iterations. The superpixels are generated by clustering

pixels based on their intensity and spatial proximity. The clustering proximity is defined as follows

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2} \qquad (2.9)$$

$$d_c = \sqrt{(l_j - l_i)^2} \qquad (2.10)$$

$$d_s = \sqrt{(x_j - x_i)^2 - (y_j - y_i)^2} \qquad (2.11)$$

where d_c is the intensity proximity, d_s is the spatial proximity and m is a constant which controls the relative weight between intensity and spatial proximity.

2.2.4 Feature Extraction

Features are any characteristics or primitives of an image that assist to distinguish one object from another object. A good feature describes the object in a meaningful manner which helps the recognition process and in the discrimination of these objects. Feature extraction is a task of extraction and generation of features to assist the task of object classification. Feature generation process computes new variables (features) from the stored values of an image or a region within an image that have high information packing properties for discrimination of different objects (Theodoridis & Koutroumbas 2003). Features can be roughly divided into: intensity features, morphological features and texture features.

2.2.4.1 Intensity Features

Pixel intensities are the simplest feature available for pattern recognition. Intensity based features are a common choice for mass detection since masses usually have an intensity higher than the surrounding tissues (Cheng et al. 2006). Contrast, average grey level (mean), standard derivation or variance, skewness and kurtosis of ROIs are some of the intensity based features (Petrick et al. 1999, Li et al. 2001, Varela et al. 2007). However, the intensity based features may fail

in classifying masses located in dense backgrounds because of the similarity in intensity values between masses and normal dense tissues.

2.2.4.2 Morphological Features

Morphological features are also called shape features. Shapes of human organs like the heart, liver, and kidney are well known, and they do not deviate much in shape from normal cases. However, the presence of disease like cancer can change the structural organization and can cause deviation from normal shape (Rangayyan 2005). Hence, analysing shape features can help in understanding/predicting the presence of any abnormality. Similarly, benign and malignant breast masses differ in shape. Masses with ill-defined and spiculated borders have a high chance of malignancy, while benign masses are associated with a circular/oval shape and smooth borders. Numerous techniques have been developed in the literature for shape analysis (Rangayyan et al. 1997, 2000, Sahiner et al. 2001, Cheng et al. 2006, Sonka et al. 2007, Oliver et al. 2010).

Shape features are based on parameters such as margin spiculation, margin sharpness, area, compactness, circularity, convexity and rectangularity of the ROI (Cheng et al. 2006, Chuand et al. 2015, Dhungel et al. 2016, Choi et al. 2016, Chokri & Farida 2017) and hence depend on the accuracy of the segmentation step. Segmentation of masses located in dense background is a challenging task due to the overlap of normal dense tissues and obscured mass boundaries (Obenauer et al. 2006, Bajger et al. 2009). In this thesis, we have presented two techniques (Chapter 6) to improve the segmentation of masses located in dense background.

2.2.4.3 Texture Features

Texture features contain information about the spatial distribution of intensity variations in an image or a selected region of an image. They are widely used in mammographic image analysis (Sahiner et al. 1998, Mudigonda et al. 2001, Choi & Ro 2012, Lin et al. 2014, Abdel-Nasser et al. 2015). As texture features are the main focus of this thesis, they have been described in detail in Section 2.1.

In addition to the above described three feature extraction methods, deep learning is seen to be gaining momentum in machine learning due to its capability to extract features at deeper levels of abstraction using multiple layers of neural

networks. Recently, a few studies have used deep learning models using Convolutional Neural Networks (CNN) for analysis of masses in mammograms (Kooi et al. 2017, Levy & Jain 2016, Dhungel et al. 2017). However, CNN features may not be suitable for mass classification in dense background due to scarcity of available data. CNNs need large training sets (typically a few thousand images) to produce accurate classification results.

2.2.5 Feature Selection

Feature selection is a statistical technique for reducing extracted features to a meaningful and compact set of features. There are many reasons for doing feature selection. One of the reasons is, not all features extracted have significant mass discrimination information. Another reason is that high feature dimensionality carries the risk of classifier overfitting and lack of generalization (Friedman 1997). Feature selection can save computation time and improve the accuracy of the classification. Feature dimensionality reduction is highly important, especially when considering scarcity of labelled data for dense breasts.

Let $X = \{x_i | i = 1, 2, \ldots n\}$ be the feature set, then feature selection process tries to find a subset $x = \{x_i | i = 1, 2, \ldots m\}$, where $m < n$, that optimizes an objective function. Traditionally, a feature selection process includes two components: a search strategy for selecting the subsets of a feature vector and an objective function to evaluate the class discrimination capability of the selected feature subset. Exponential search techniques like exhaustive search, sequential search techniques like sequential forward and backward search, and random search techniques like genetic algorithms are some of the most commonly used search strategies.

The evaluation or the objective functions can be broadly categorized into two groups: wrapper methods and filter methods. Wrapper methods evaluate feature subsets by training a model on it and evaluating their predictive performance through statistical resampling or cross validation. Filter methods do not use any training model, instead evaluate the feature subset by analysing the general characteristics of the training data such as intra-class and inter-class variance.

An exhaustive search examines all possible subsets of the feature vector in order to find the best subset satisfying the objective function. This method guarantees finding an optimal feature subset. With n features generated there are $2^n - 1$ feature combinations. This makes exhaustive search computationally expensive and not practical for larger values of n. However if we limit the maximum

number of features to d, the exhaustive search is reduced to $\binom{n}{d}$ combinations and $\binom{n}{d} < 2^n - 1$. In this thesis, we have used exhaustive search feature selection.

Sequential search is a computationally efficient search technique, but a sub-optimal feature selection technique as it does not examines all possible subsets of the feature vector. In addition, sequential methods have a tendency to fall in local minima. Sequential feature selection methods can be broadly classified into four categories: sequential forward selection, sequential backward selection, sequential floating feature selection and bidirectional feature selection. The sequential forward selection method starts with an empty set and features are added to it. If the feature increases the performance of the classifier substantially, it can be added to a semi-optimal feature set. If the contribution of the feature is minimal, the feature can be ignored. This process is continued until the addition of further features does not improve the objective function. The sequential backward selection method starts with the complete set of attributes. At every stage, the procedure removes the worst attribute from the set leading to a reduced set. Two sequential floating methods are sequential floating forward selection and sequential floating backward selection. The forward selection method starts with an empty set and makes, after each forward selection, several backward steps if the objective function improves. In contrast, the backward selection method starts with the complete set and makes, after each backward selection, several forward steps if the objective function improves. Bidirectional feature selection incorporates the sequential forward and backward feature selections. The procedure begins with sequential forward selection from an empty set, simultaneously with sequential backward selection from a full set. The procedure converges to a semi-optimal feature set when added features are not removed and removed features are not added again.

Random search, like genetic algorithms, incorporates randomness into their search to minimise the problem of falling into local minima. Genetic algorithms (GA) are based on the process of natural selection that evolves a population over time to better adapt to the surroundings. GA starts from a random population of individuals called candidate solutions and associated properties for each candidate called chromosome. Each chromosome is represented in binary strings of 0s and 1s. The number of digits in the binary string represents the total number of features, and 1 and 0 represents the inclusion and exclusion of particular features in the classification model. A fitness function is used to evaluate the performance of each individual in solving the problem and the fittest individuals are selected for producing the next generation individuals. The selected individual's

chromosomes are mutated and altered using two genetic operators (crossover and mutation) to produce new generation individuals. The algorithm usually terminates when the maximum number of allowed generations have been produced or when two subsequent generations no longer differ significantly from each other.

2.2.6 Classification

After the feature selection process, the selected features are given as input to a classifier to categorize regions identified as suspicious into normal tissue or malignant mass. Supervised and unsupervised are the two main classification techniques. If a set of training data is available and the classifier is designed by exploiting this prior information, then it is known as supervised classification while in unsupervised classification, no prior training information is available. Fisher Linear Discriminant Analysis, Artificial Neutral Networks, Decision Tree, Random Forest and Support Vector Machine are some of the popular supervised classification techniques. In this thesis we have used Fisher Linear Discriminant Analysis and Support Vector Machine classifiers for the experiments.

2.2.6.1 Fisher Linear Discriminant Analysis

Fisher Linear Discriminant Analysis (LDA) is one of the most commonly used mass classification methods. It was originally developed by Fisher (1936). Let $x = [x_1, x_2, \ldots, x_d]$ be the feature vector and $w = [w_1, w_2, \ldots, w_d]$ be the corresponding weight vector, then the linear discriminant function is defined as follows:

$$J(x) = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d = w^T x. \tag{2.12}$$

The main idea of LDA classification is to find a value for w that maximizes the value of Z. This provides a ratio between the inter-class and intra-class variances, in order to increase the separation between two classes. The Fisher criterion Z is defined as follows (Ganesan et al. 2013a):

$$Z = \frac{|w^T(\mu_1 - \mu_2)|^2}{w^T S w}, \tag{2.13}$$

where μ_1 and μ_2 are the two class observations means and S is the within-class

scatter matrix which is defined as follows:

$$S = \sum_{i=1,2} \sum_{j \in class_i} (j - \mu_i)(j - \mu_i)^T. \tag{2.14}$$

After finding the value of w based on the training set, the new observation is classified as abnormal or class 1, if $J(x)$ is positive, otherwise it is normal or class 2.

2.2.6.2   Support Vector Machine

Support Vector Machine (SVM) is another widely used mass classification method, originally developed by Vapnik & Lerner (1963). The idea of SVM mass classification is to design a hyperplane or a set of hyperplanes that can be used for classifying the training data into two classes (normal, abnormal). An SVM hyperplane can be represented in the form of $f(x)$ as given below.

$$f(x) = \omega^T x + \omega_0 = 0. \tag{2.15}$$

The hyperplane which can separate the training data into two classes is called a separating hyperplane and optimal separating hyperplane will be the hyperplane that separates the classes with maximal distance margin from both classes. Let $\omega^T x + \omega_0 \geq 1$ for all $x \in class_1$ and $\omega^T x + \omega_0 \leq -1$ for all $x \in class_2$ be two parallel hyperplanes. Let $f(x)$ be the separating hyperplane, then the margin is the distance from any point that lies on either of the two hyperplanes to the separating hyperplane. It is calculated as follows:

$$d = \frac{\mid f(x) \mid}{||w||} = \frac{1}{||w||}. \tag{2.16}$$

To maximize the distance, the term ω needs to be minimized. Minimizing ω is a non linear optimization task, solved by the Karush-Kuhn-Tucker (KKT) conditions, using Langrange multipliers λ_i.

$$w = \sum_{i=1}^{N} \lambda_i y_i x_i, \tag{2.17}$$

$$\sum_{i=1}^{N} \lambda_i y_i = 0. \tag{2.18}$$

The SVM algorithm was generalized to non linear classifier by applying kernel function where the feature sample from a lower dimensional space is mapped to a higher dimensional space for making the classification easier (Guyon et al. 1993). Polynomials, radial basis function, and hyperbolic tangent are some of the most commonly used kernels (Ganesan et al. 2013b).

2.2.6.3 Artificial Neutral Networks

Artificial Neutral Networks (ANNs) are the collection of mathematical models that imitate the properties of biological nervous system and the functions of adaptive biological learning. They are made of many processing elements that are highly interconnected together with the weighted links that are similar to the synapses. The advantage of ANNs is their capability of self-learning and because of that, they are often suitable to solve the problems that are too complex to use the conventional techniques, or hard to find algorithmic solutions (Cheng et al. 2006). In the basic form, the perceptron learns a linear decision function that divides two linearly separable training sets. In ANN a perceptron is a mathematical model of a biological neuron. The response of this basic model based on weighted sum of its inputs can be represented as given below.

$$d(x) = \sum_{i=1}^{n} w_i x_i + w_{n+1} \tag{2.19}$$

where the coefficients $w_i, i = 1, 2, \ldots, n, n+1$ called weights, modify the inputs before they are summed and fed into the threshold element. Here, the function that maps the output of the summing junction into the final output of the device is called activation function. When $d(x) > 0$, the threshold element causes the output of the perceptron to be $+1$, indicating that pattern x belongs to class $C1$ and $C2$ if $d(x) < 0$.

2.2.6.4 Decision Tree

Decision tree is a multistage decision making processes where instead of using a complete set of features jointly to make a decision, different subsets of features are used at different levels of the tree (Quinlan 1986). Decisions on choosing the classes are done in a sequential manner in this technique. Decision trees start from one parent node and continue splitting until a final result is obtained. Splitting criteria are set for each of the nodes and adherence to the splitting criterion is

necessary for node splits. With each split of the node, the node is declared as a leaf and a particular class label is to be given to that node. Majority rule can be used to label the node as stated in Equation 2.20.

$$k = argmax_i P(w_i|t) \tag{2.20}$$

where the rule states that a leaf t is assigned to a class where the majority of the vectors in the node tree belongs.

2.2.6.5 Random Forest

Random forest (RF) is a popular ensemble classification method, originally developed by Ho (1995). Ensemble algorithms combine multiple learning algorithms to achieve better performance than could be achieved from any of the single learning algorithms alone. RF builds up hundreds to thousands of decision trees and obtain vote from each tree and then classifies using majority vote. The main steps of RF algorithm are as follows:

- If N is the number of samples in the training set, then randomly select N samples but with replacement for growing the trees.

- At each node of the tree, randomly select m variables out of M ($m < M$) for splitting. Grow each tree to the largest extend possible.

- Predict new data by aggregating information from all trees. Let $C_f(x)$ be the class prediction of the fth random forest tree out of F. Then $C_{rf}^F(x) = $ *majority vote* $\{C_f(x)\}_1^F$ (Friedman et al. 2001).

2.2.7 Validation

Classifiers are affected by noise and outliers present in the dataset. Some popular techniques for evaluation of the classification models are k-fold cross validation and leave one out cross validation.

In k-fold cross validation, the original dataset is randomly partitioned into k subsets of equal size. For example, if the dataset has 100 instances, then 10 datasets are created with 10 instances each. Each time a classifier is tested, k-1 subsets are considered as the training set and the remaining set is treated as the test dataset. The process is then repeated for k trials. The overall performance is

the average of all k trials. The commonly used k-fold cross validation are 5-fold cross validation and 10-fold cross validation where the value of k is 5 and 10, respectively. To ensure stable classification results, multiple independent runs of 5-fold or 10-fold cross validation can be performed. In this thesis, we have used 10 independent runs of 5-fold cross validation in Chapters 3, 4 and 5.

Leave one out cross validation is a special form of k-fold cross validation, where the value of k is the same as the number of instances in the dataset. That is every instance is considered to be a dataset. For example, if the dataset has 100 instances, then 100 datasets are created with 1 instance each. The classifier is trained with 99 instances and tested with one instance. Leave one out cross validation is unbiased and is highly preferred when the dataset is very small. In this thesis, we have used leave one out cross validation in Chapter 7.

2.2.8 Evaluation

Receiver operating characteristics (ROC) curve is a widely used performance measure for medical diagnostic classification. ROC curve is a plot of classifiers true positive rate (sensitivity) versus false positive rate (1-specificity) (Bradley 1997). It is used to assess the predictive power of a classifier by using Area Under the ROC curve (AUC). Sensitivity (true positive rate) is defined as the proportion of correctly detected positive instances (correctly detected cancer ROIs) over total number of positive instances (total number of cancerous ROIs). Similarly, specificity (true negative rate) is defined as the proportion of correctly detected negatives instances (correctly detected normal ROIs) over total number of negative instances (total number of normal ROIs). A false positive is where the CAD system wrongly classify an ROI as positive (cancer). The value of AUC is 1.0 when the classification is perfect which means that sensitivity is 100% and false positive rate is 0%. ROC analysis is mainly used for mass classification while for mass detection (localisation of the tumor), free-response receiver operating characteristics (FROC) is used. An FROC curve is obtained by plotting sensitivity (correctly detected masses) on vertical axis and average number of false positive detections per image on the horizontal axis. In this thesis, we have used AUC, sensitivity, specificity for mass classification and FROC for localisation of the tumor.

2.3 Database

A common description of the datasets used in this study (Chapters 3-7) is presented here to avoid repetition in subsequent chapters. The data used in this research are taken from two databases. The publicly available Digital Database for Screening Mammography (DDSM) (Heath et al. 2001) and BreastScreen SA (BSSA), a local screening archive in Adelaide, South Australia.

The dense background in this study refers to density of the immediate background environment of the mass and not the overall density of the breast as used in the BI-RADS. Figure 2.1 illustrates the difference. A mass may be located in a non-dense area of a breast despite high BI-RADS density category and similarly low BI-RADS density score does not exclude masses located in dense tissues.

(a)   (b)
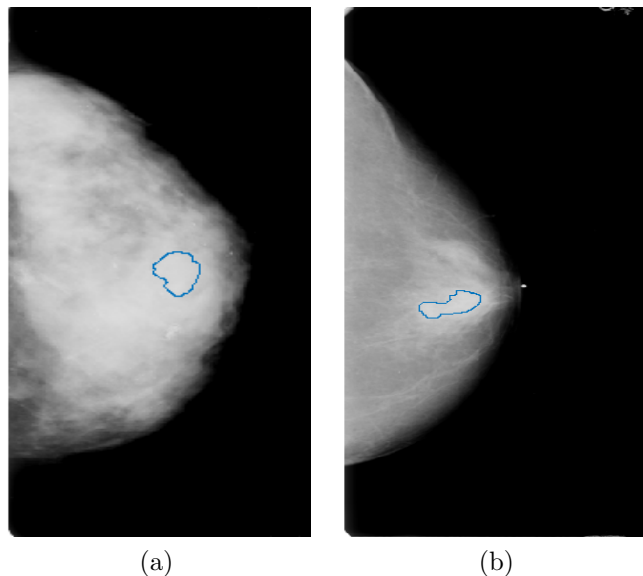
Figure 2.1: Examples of mammograms showing masses in dense backgrounds. (a) BI-RADS category IV (Glandularity > 75%) and (b) BI-RADS category II (Glandularity 25 − 50%).

Mass annotations were provided in the DDSM database; however, the annotation contours were often very generous. Similarly, mass annotations in the BSSA database were provided using a rectangular box indicating only approximate location of the mass. For this reason, the core mass contours for both databases were manually delineated for this study by the author under the guidance of an experienced radiologist (more than 20 years) in mammography. All the annotations were reviewed and validated by the radiologist. These core mass contours were used in experiments in Chapters 3, 4, 5, 6 and 7. ImageJ software package (Schneider et al. 2012) was used to draw the contours.

2.3.1 Public Databases

The Mammographic Image Analysis Society (MIAS) (Suckling et al. 1994), Digital Database for Screening Mammography (DDSM) and INbreast (Moreira et al. 2012) are three well-known and well-tested public databases available to researchers. The MIAS database is from an organization of UK research groups while DDSM is a collaborative effort between Massachusetts General Hospital, Sandia National Laboratories and the University of South Florida Computer Science and Engineering Department. The INbreast database is from a breast center located in a university hospital, Portugal. Both MIAS and DDSM contain digitized film mammograms whereas INbreast contains full field digital mammograms (FFDM). The MIAS and INbreast databases were explored as a part of this study, but not included for further research due to an insufficient number of mammograms with masses in dense background (less than 10), which is the main focus of this study. Hence, DDSM is the only publicly available database used in this study.

The DDSM database contains mammographic images from approximately 2500 women. For each woman, four images were taken (CC and MLO views for each breast). The mammograms were acquired using four scanners: Lumisys 200 Laser (50 μm pixel size, 12 bits), DBA M2100 ImageClear (42 μm pixel size, 16 bits), Howtek 960 (43.5 μm pixel size, 12 bits) and Howtek Multi-Rad850 (43.5 μm pixel size, 12 bits). We selected all the malignant cases satisfying our criteria (malignant masses located in dense background). This resulted in 41 cases (41 mammograms from 41 different women). In addition, 41 malignant mammograms (41 cases) with mass in nondense background and 52 normal dense mammograms (26 cases - either CC or MLO view for each breast) were identified and employed in this study. Table 2.3 shows the statistics. For mass region, a subtlety rating indicating the difficulty level of detecting mass is provided. A lower rating indicates higher level of difficulty in detection. Figures 2.2 and 2.3 show the mass subtlety and distributions of mass size for masses in dense background dataset and masses in nondense background dataset respectively.

2.3.2 BreastScreen SA

The Computer Aided Screening Mammography (CASM) group at Flinders University has a long standing agreement with BreastScreen SA to digitize screening film mammograms in their archives. The CASM group has collected over one

Table 2.3: Statistics of DDSM dataset.

	Number of cases	Number of mammograms	Number of masses
Mass in dense background (malignant)	41	41	41
Mass in nondense background (malignant)	41	41	41
Normal	26	52	-
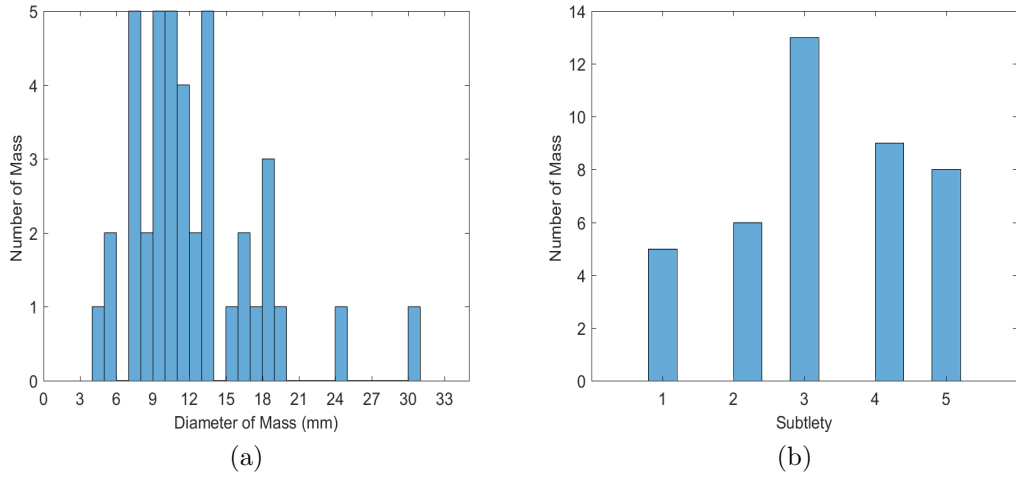
Figure 2.2: Information of DDSM masses in dense background (a) Diameter of masses in millimeter (b) subtlety of masses.
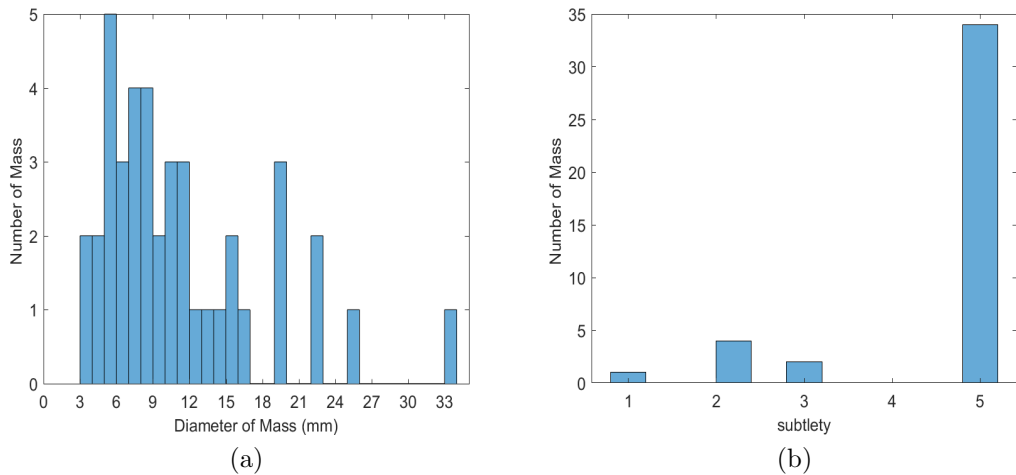
Figure 2.3: Information of DDSM masses in nondense background (a) Diameter of masses in millimeter (b) subtlety of masses.

thousand de-personalized screening mammograms, but the metadata (details of the mammogram - cancerous/normal etc) was not available in a digital form. The author collated the associated metadata and prepared the BSSA database

Table 2.4: Statistics of BSSA dataset.

	Number of cases	Number of mammograms	Number of masses
Mass in dense background (malignant)	29	29	29
Normal	39	39	-
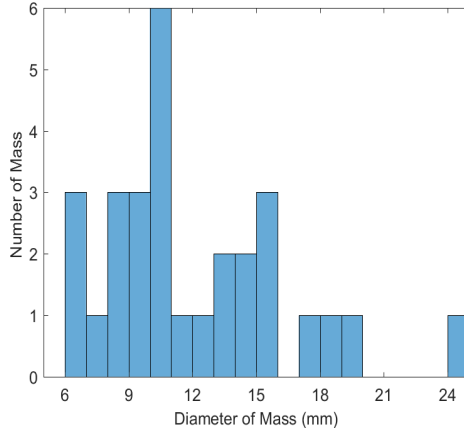
Figure 2.4: Diameter of BSSA masses in millimeter.

for the purpose of this study. All the malignant mammograms having masses located in dense background (29 cases) and 39 normal dense mammograms (39 cases) were selected for this study. Table 2.4 shows the statistics. The collection of images and information corresponding to one woman is referred as a "case". One mammogram (either left or right breast taken in either CC or MLO view) per case is selected for this study.

The mammograms in BSSA were acquired using a Vidar Diagnostic Pro Advantage digitiser (48 μm pixel size, 12 bit). Figure 2.4 shows the distributions of mass size. The dataset does not contains the information about the difficulty level of detecting the mass.

Chapter 3

Structured Micropattern Analysis

The main focus of this thesis is to find efficient texture descriptors for masses localized in dense background of mammograms. In this chapter, a novel approach is presented for classification of mammographic dense regions as malignant masses or normal dense regions using structured micropatterns generated from LBP. Section 3.1 provides a brief overview of various methods used to describe texture structures. Section 3.2 describes the LBP technique which is used in this chapter to generate structured micropatterns and Section 3.3 provides details of the structured micropatterns. Section 3.4 describes datasets used for conducting experiments in this chapter and Section 3.5 presents the methodology. Experimental set up and results are presented in Section 3.6 and Section 3.7, respectively. Section 3.8 provides a direct comparison of proposed method with a recent state-of-the-art mass classification method. Discussion and conclusion are provided in Section 3.9 and Section 3.10, respectively.

3.1 Introduction

Texture features have proven to be useful in discriminating different classes in many applications of computer image analysis. Studies (Varma & Zisserman 2003, Li et al. 2014) have shown that local features are more effective than global ones for texture classification. The local texture features are extracted based on the local neighborhood of image pixels, describing the spatial intensity distribution in a pixel's neighborhood. The literature presents many approaches for design of micropatterns to describe the local spatial context of the image (Lowe

1999, Yang et al. 2005, Wei et al. 2007, Ravela & Manmatha 1997, Hadjidemetriou et al. 2004). Lowe (1999) introduced scale invariant feature transformation to calculate the regional characteristics of micropatterns used for object recognition. Yang et al. (2005) used Markov Random Field method to design adaptive micropatterns for face identification. The micropattern has also been extended to Gabor filters (Wei et al. 2007), difference of Gaussians (Ravela & Manmatha 1997) and with multiresolution histogram (Hadjidemetriou et al. 2004) to encode the structure information.

Local Binary Pattern (LBP) (Ojala et al. 1996, 2002) is another popular technique used to describe micropatterns. It is one of the most discriminative and computationally simple local texture descriptors and has been investigated for about two decades (Ojala et al. 1996, Pietikainen & Zhao 2015). Due to its ease of implementation and ability to describe fine texture details, LBP has gained significant attention and has been applied in various fields such as facial image analysis (Zhang et al. 2005, Tan & Triggs 2010), biometrics (Bai & Hatzinakos 2010), medical image analysis (Llado et al. 2009, Choi & Ro 2012), motion analysis (Heikkila & Pietikainen 2006) and content based retrieval from image or video databases (Liao & Chen 2002). Examples of research using LBP in medical fields include search and retrieval methods for finding relevant slices in brain MRIs (Unay & Ekin 2008), textural features extraction from thyroid ultrasounds (Keramidas et al. 2008), automated cell phenotype image classification (Nanni & Lumini 2008) and mammogram breast density classification (George et al. 2018).

In the field of mammographic mass classification, Llado et al. (2009) used LBP to represent textural properties of masses for reduction of false positives in computer-aided detection of breast masses in mammograms. The study followed the original idea of the LBP operator proposed by Ojala et al. (1996, 2002), where parts of an image (Regions of Interests (ROIs)) were used to compute the LBP histograms which were later concatenated to form a high dimensional (more than 500) feature vector. Choi & Ro (2012) extracted LBP patterns from two specific regions of the ROI (core and margin regions) for classifying ROIs as either breast cancer or normal breast tissues. Again, the approach was based on histogram concatenation which yielded a high-dimensional feature space (255). It can be noted that the histogram concatenation approach does not recognize any specific patterns structure in the feature space, thus, it naturally leads to a very large dimension.

In this chapter, a novel method for dense breast mass classification utilizing

specific structures of LBP generated patterns, the structured micropatterns, is proposed. The typical histogram concatenation approach of extracting features from LBP is replaced by features based on combinatorial properties of the structured micropatterns. Using combinatorial properties of the structured micropatterns, a small set of features was extracted, which enabled successful classification of cancerous dense ROIs.

3.2   LBP Overview

A preliminary study of encoding information in an image by mapping a local neighborhood surrounding a pixel was first presented by He & Wang (1990). In their approach, the 3×3 neighborhood of a central pixel is assigned a value of 0, 1 or 2 according to this rule:

$$
e_i = \begin{cases} 0, & \text{if} \quad I_i < I_c, \\ 1, & \text{if} \quad I_i = I_c, \\ 2, & \text{if} \quad I_i > I_c, \end{cases} \tag{3.1}
$$

where I_c is the intensity of the central pixel and $I_i \{i = 1, 2, \ldots 8\}$ is the intensity value of a neighboring pixel. The central pixel, or a texture unit (T), is represented as follows:

$$
T = \sum_{i=1}^{8} 3^{i-1} e_i. \tag{3.2}
$$

This representation produces $3^8 = 6561$ possible texture units for describing the three level patterns in eight directions. Later Ojala et al. (1996) introduced a two level version of the Wang and He method, known as Local Binary Pattern. In Ojala's approach, there are only $2^8 = 256$ possible texture units instead of 6561. The 3×3 neighborhood is assigned a value of 0 or 1 (see Figure 3.1 for an illustration) instead of three possible values considered in (He & Wang 1990).

The LBP$_{P,R}$ for a pixel (x_c, y_c) is calculated by comparing this (central) pixel intensity value g_c with intensities of its P neighboring pixels $(g_0, \cdots, g_{P-1})$ on a
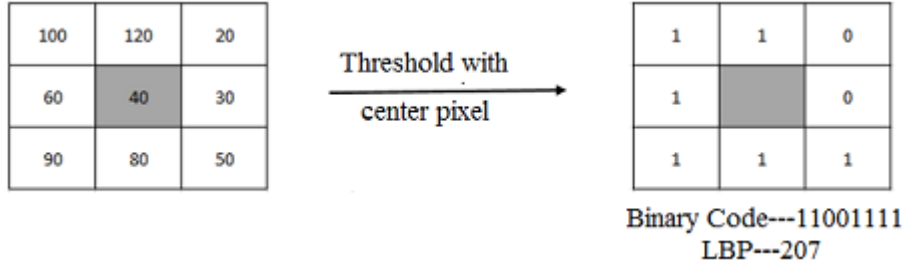
Figure 3.1: Example of basic LBP operator for 3×3 grid neighborhood.

circle of radius R. It is computed as follows:

$$LBP(x_c, y_c) = \sum_{i=0}^{P-1} A(g_i - g_c)2^i \qquad (3.3)$$

where $A(g_i - g_c)$ is 1 if $g_i \geqslant g_c$ and 0, otherwise. If the coordinates of the center pixel (x_c, y_c) are (0,0), then the coordinates of neighboring pixels (x_i, y_i) are given by $x_i = Rcos(2\pi i/P)$ and $y_i = -Rsin(2\pi i/P)$. The pixel intensities of the neighbors which do not fall exactly on the center of the pixels are calculated by interpolation. Figure 3.2 shows a circular neighborhood LBP approach with P=8 and R=1, 2, 3 and 4. The histogram of LBP labels is often used in the literature as a texture descriptor.
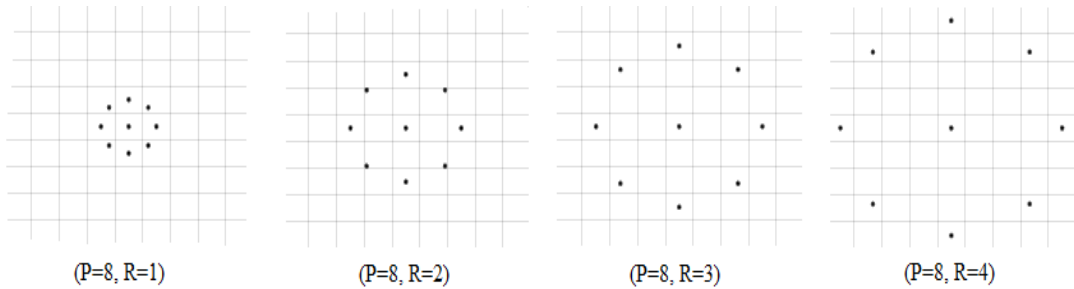
Figure 3.2: Example of a circularly symmetric LBP neighborhood around a central pixel with P=8 and R=1, 2, 3 and 4.

The LBP operator is made rotation invariant by performing P bit-wise shifts, where P is the number of neighbors, and selecting the smallest value of the binary pattern (Pietikäinen et al. 2000, Ojala et al. 2002). An LBP pattern is considered 'uniform' if the number of transitions between 0 and 1 in the sequence is at most two, when viewed as a circular string.

Even though LBP is a simple and efficient texture descriptor method, the major issue with the LBP histogram based texture descriptor is that the feature

space increases exponentially with the number of LBP neighbors. For example, a uniformly sampled circular neighbourhood of P pixels at radius R ($LBP_{P,R}$) has 2^P histogram bins. For P = 8, LBP produces 256 histogram bins which represents 256 features, whereas one with 24 neighbors produces 16,777,216 features. In the typical LBP approach to find features, an image is subdivided into rectangular blocks and the LBP histogram for each block is obtained, which are later concatenated to form a high dimensional feature vector. This may result in increased computational complexity, redundant information and poor generalization properties of the classifiers.

3.3   Structured Micropatterns

LBP produces 36 rotationally invariant LBP patterns for 8 neighbourhood. Among these 36, nine are categorized as uniform. The nine uniform rotationally invariant local binary patterns are: p_0 (00000000), p_1 (00000001), p_3 (00000011), p_7 (00000111), p_{15} (00001111), p_{31} (00011111), p_{63} (00111111), p_{127} (01111111) and p_{255} (11111111). Figure 3.3 illustrates the nine patterns. Ojala et al. (2002) stated that these nine rotational invariant uniform patterns are fundamental properties of texture, providing the majority of patterns, over 90% of $LBP_{8,1}$ patterns. These patterns can describe micro-structures like edges, lines, spots and flat areas that refer back to Julesz Textons which are the basic elements of pre-attentive human perception (Julesz 1981). Another advantage of these nine rotational invariant uniform patterns is their statistical robustness (Pietikainen et al. 2011). They are less prone to noise when compared to non uniform patterns.
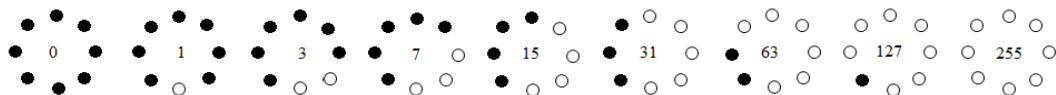
Figure 3.3: Nine uniform rotational invariant LBP patterns. Black and white circles represents the bit values of 0 and 1, respectively, in the 8 bit LBP pattern and the numbers inside represents the LBP label for that particular pattern. The first pattern is called *spike* and the last one is named *pit*.

The mass ROI classification capability of these 9 structured micropatterns (nine uniform rotational invariant patterns) generated using LBP technique are analyzed in this chapter. Female breast is made of different types of parenchymal patterns and the presence of masses can change the composition of these patterns (Midya & Chakraborty 2015). Normal dense regions are mostly homoge-

neous and have regular and repetitive patterns, whereas cancer/masses typically have random patterns due to their heterogeneous nature (Rabidas et al. 2018b). From these observations, it can be hypothesized that the density/proportion of the structured micropatterns found in a mass region may differ from those found in a normal dense region. Hence these local geometrical features could be used to distinguish the mass from a normal dense region. This study of structured micropatterns was motivated by the above heuristic.

3.4   Data

The data used in the experiments were taken from both DDSM and BSSA databases (Section 2.3). As mentioned previously, the core mass contours were manually delineated by the author under the guidance of an experienced radiologist in mammography. These core mass contours were used to generate the breast mass ROIs. Using MATLAB the smallest rectangular region containing the core mass contour was extracted. We call it mass ROI. Normal dense ROIs were selected manually from dense regions of a healthy dense breast. Figure 3.4 shows an example of a normal dense ROI cropped from a dense healthy breast. There was no overlap between the ROIs selected from the same mammogram. Figure 3.5 shows an example of mass ROIs (a and b) and healthy dense region ROIs (c and d). The figure reveals that, it is a very challenging task to identify which ROI contains the cancer or not as both mass ROI and normal ROI looks very similar.

As described previously, all the malignant mammograms from the DDSM database satisfying the criteria (malignant masses located in dense background) were selected for this study. This resulted in 41 malignant mass ROIs (41 mammograms from 41 different women and one mass per mammogram). Normal dense ROIs were selected from 52 dense breast mammograms (52 mammograms from 26 different women and 5 ROIs from each mammogram) which resulted in 260 normal ROIs. This makes a total of 301 ROIs selected from DDSM for this study. From the BSSA database, all the available malignant mammograms satisfying the criteria (malignant masses located in dense background) were selected. This resulted in 29 malignant mass ROIs (29 mammograms from 29 different women and one mass per mammogram). The normal dense ROIs were selected from 39 dense breast mammograms (39 mammograms from 39 different women and 5 ROIs from each mammogram) which resulted in 195 normal ROIs. This
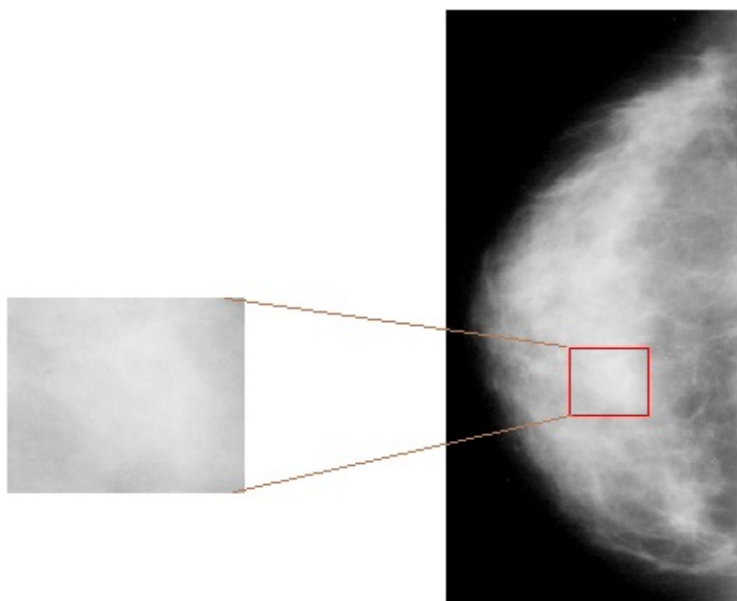
Figure 3.4: An example of a dense healthy breast and an ROI cropped from dense region

(a) Mass ROI   (b) Mass ROI   (c) Normal ROI   (d) Normal ROI
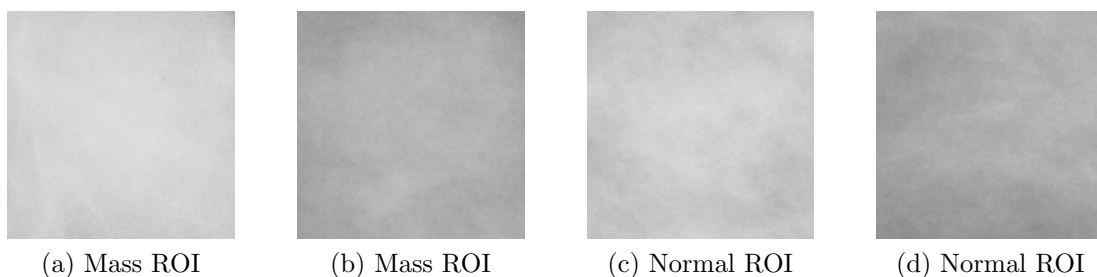
Figure 3.5: Examples of Mass and Normal ROIs.

makes a total of 224 ROIs selected from BSSA for this study. Table 3.1 shows the statistics of the datset. The same ROIs (301 from DDSM and 224 from BSSA) are used in next two chapters (Chapters 4 and 5).

Table 3.1: Dataset Information.

	DDSM	BSSA
Mass ROIs	41	29
Normal ROIs	260	195

For texture analysis and classification (Chapters 3, 4 and 5) experiments, 260 DDSM normal dense ROIs were divided into five subsets (five ROIs from each of 52 healthy dense breast mammograms). This was then combined with 41 malignant dense mass ROIs. This resulted in five balanced datasets with 93 ROIs

in each (52 normal and 41 malignant mass ROIs). Figure 3.6 shows the dataset subdivision procedure. The same procedure was repeated for the BSSA database to get five balanced datasets of 68 ROIs in each (39 normal and 29 malignant mass ROIs). Figure 3.7 shows the BSSA dataset subdivision procedure. For feature selection in this chapter and Chapters 4 and 5, only the DDSM subset (Dataset 1 shown in Figure 3.6) was used and for evaluation extended DDSM (all 5 datasets) were used. All 5 BSSA datasets were used in the evaluation stage. BSSA datasets served as unseen data for the classification models, as they were not used at any stage of the feature selection process.
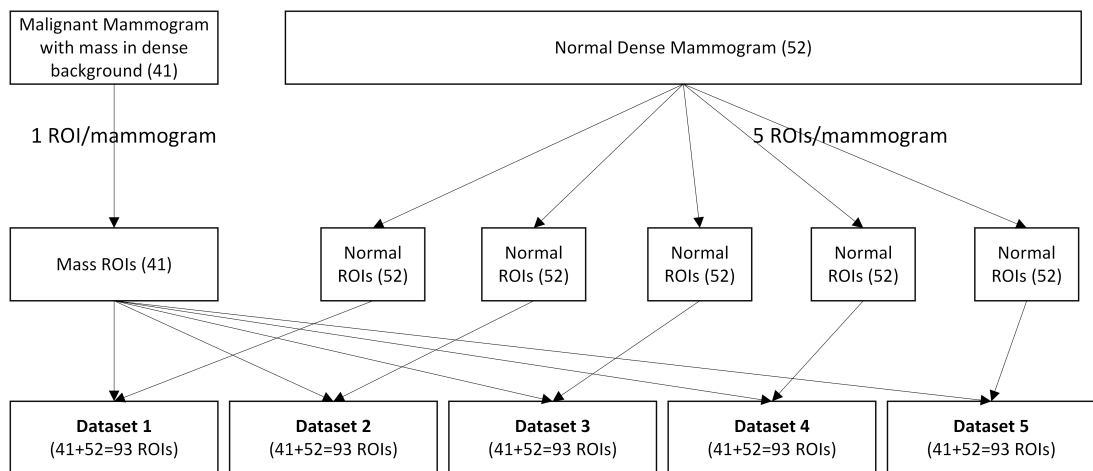
Figure 3.6: DDSM dataset subdivisions: 41 mass ROIs from 41 mammograms with masses located in dense background and 52 normal dense ROIs from 52 normal dense breast mammograms combined to form each datasets. The mass ROIs were the same in all five datasets as shown in figure.

3.5   Methodology

The mass classification approach consists of five steps: denoising, ROI extraction, structured micropatterns extraction from ROI, feature generation and classification using the features generated from structured micropatterns. Figure 3.8 shows the flowchart for the mass classification scheme.

3.5.1   Denoising

Mammographic images contain noise and LBP is susceptible to noise. Hence all images from DDSM and BSSA were denoised, to alleviate the influence on the classification process, using a 3×3 median filter before ROIs extraction.
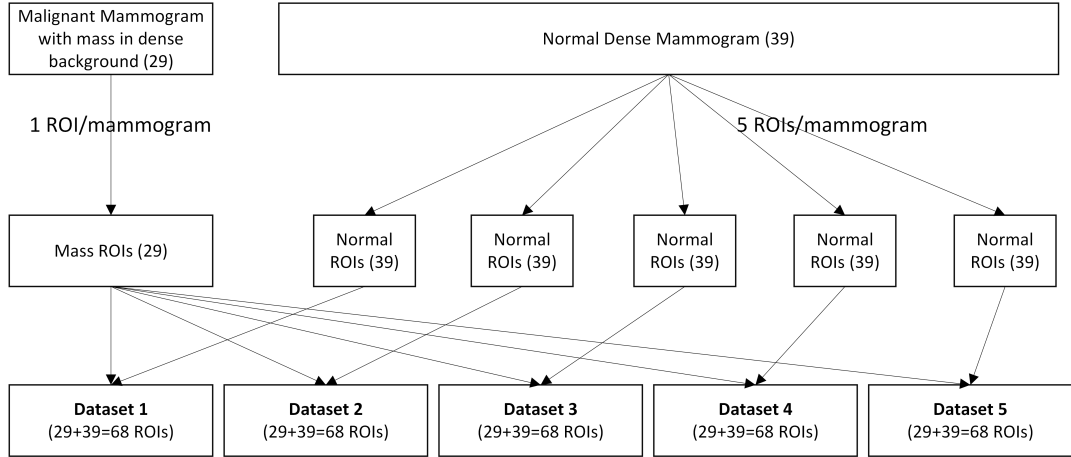
Figure 3.7: BSSA dataset subdivisions: 29 mass ROIs from 29 mammograms with masses located in dense background and 39 normal dense ROIs from 39 normal dense breast mammograms combined to form each datasets. The mass ROIs were the same in all five datasets as shown in figure.
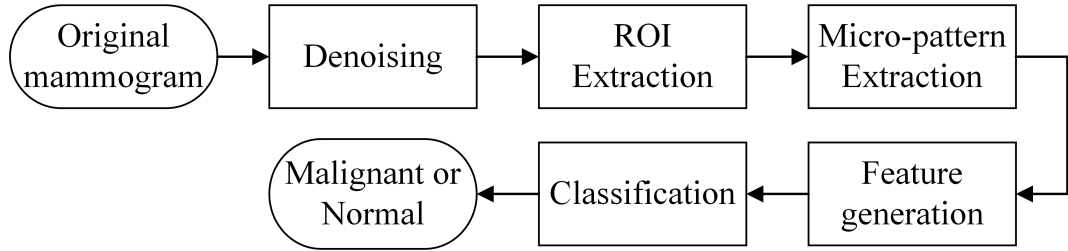
Figure 3.8: Flowchart for the mass classification algorithm

3.5.2   Feature Generation

The LBP technique proposed by Ojala et al. (2002) (as described in Section 3.2) was used to build the LBP labeled image. For the experiments, the number of neighbors P was selected to be 8 and the radius $R \in \{1, 2, 3, 4\}$. The number of occurrences of each of the nine rotationally invariant LBP patterns (structured micro patterns) per unit area is used as a feature for the purpose of discriminating mass ROI and normal dense ROI. That is, for each ROI the features are defined as follows:

$$f_i = \frac{|pattern_i|}{|ROI|}, \tag{3.4}$$

where $|.|$ stands for the number of elements in the set, $i = 0, 1, 3, 7, 15, 31, 63, 127, 255$ and $pattern_i$ is the ith structured micropattern. Using notation introduced in Section 3.3, the features corresponding to the 9 structured micropatterns: p_0,

49

p_1, p_3, p_7, p_{15}, p_{31}, p_{63}, p_{127}, p_{255} are: f_0, f_1, f_3, f_7, f_{15}, f_{31}, f_{63}, f_{127}, f_{255}. This results in 9 features for each of the 4 levels. The features generated are independent of ROI size since it was normalized with the number of pixels in each ROI.

3.5.3 Feature Selection

All feature combinations generated from the nine structured micropatterns may not contain significant mass discrimination information. Hence, these nine structured patterns were analysed individually and in combination for their capability for classifying dense mass ROI. An initial study was conducted with 8 neighbourhood and radius = 1. Exhaustive search was used for feature selection. This resulted in a total of 511 feature combinations ($2^9 - 1 = 511$). The dataset used for feature selection was a subset of the DDSM database (dataset1 (93 ROIs)- see Section 3.4). The effectiveness of different features extracted from structured micropatterns for the separation of cancerous and normal regions was evaluated using LDA. Ten independent runs of 5 fold cross validation were performed and the performance of each selection was analysed using AUC score.

While analysing 511 feature combinations, the $\{f_1, f_3, f_7, f_{31}, f_{255}\}$ feature combination achieved the highest AUC of 0.972. Pairwise T-test was conducted to see whether this highest score was significantly different from others at the $p = 0.05$ level. All feature combinations of 1, 2, 3, 4 and 5 features (in total 381 feature combinations) were tested one at a time against the highest scored feature set $\{f_1, f_3, f_7, f_{31}, f_{255}\}$. Only feature sets with up to five dimensions were tested, as the classifier performance was not increased with further increase in the feature dimension. From the T-test results it was found that the highest scored feature set $\{f_1, f_3, f_7, f_{31}, f_{255}\}$ was significantly better than others except for 81 (out of 381) feature combinations. Out of these 81 feature combinations, 5 feature sets $\{f_0, f_1, f_{255}\}$, $\{f_0, f_3, f_{255}\}$, $\{f_0, f_7, f_{255}\}$, $\{f_1, f_3, f_{128}\}$, and $\{f_1, f_{31}, f_{255}\}$ had lower dimensionality than the highest performing feature set $\{f_1, f_3, f_7, f_{31}, f_{255}\}$. Hence these 5 feature sets, not statistically different in performance than the highest scored feature set, were selected for further analysis (see Table 3.2).

3.6 Experimental Setup

The ROI classification power of the selected features was evaluated on the extended DDSM database (301 ROIs - all 5 datasets) and unseen BSSA database

Table 3.2: DDSM classification results from 10 runs of 5-fold cross validation for dataset 1 (50 values) using LDA. AUC is the mean AUC scores $\pm$ standard deviation.

Feature Set	AUC	p-value
$\{ f_1, f_3, f_7, f_{31}, f_{255} \}$	**0.972 $\pm$ 0.005**	-
$\{ f_1, f_{31}, f_{255} \}$	0.967 $\pm$ 0.006	0.576
$\{ f_1, f_3, f_{128} \}$	0.957 $\pm$ 0.006	0.096
$\{ f_0, f_7, f_{255} \}$	0.956 $\pm$ 0.004	0.229
$\{ f_0, f_1, f_{255} \}$	0.952 $\pm$ 0.005	0.084
$\{ f_0, f_3, f_{255} \}$	0.949 $\pm$ 0.007	0.051

(224 ROIs - all 5 datasets). See Section 3.4 for details. In addition to radius 1, radius 2, 3 and 4 and their combinations were analysed. LDA and SVM with linear kernel (SVM-L) were used to validate the classification power of selected features. These classifiers were chosen because of their general popularity in machine learning (Hastie et al. 2009). The MATLAB code for the classifiers (fitcdiscr for LDA and fitcsvm for SVM) were taken from the Statistics and Machine Learning Toolbox of MATLAB 2015b package with default parameter values. To ensure stable classification results, 10 independent runs of five-fold cross validation was applied for each of the 5 datasets. The results reported are averages of these 10 runs over 5 datasets (for both DDSM and BSSA). In addition to AUC score, sensitivity and specificity were reported.

3.7 Results

The five feature sets identified in Section 3.5.3 were applied to the unseen BSSA data using LDA and their highest AUC scores (out of all radius combinations) are shown in Table 3.3. Out of these five feature sets, set $\{f_0, f_1, f_{255}\}$ achieved the highest AUC score of 0.885. As it is shown in the table, two other feature sets ($\{f_0, f_3, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$) also performed at a similar level. Hence, we have presented the detailed results for these three feature sets $\{f_0, f_1, f_{255}\}$, $\{f_0, f_3, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$ in Table 3.4 and Table 3.5 for LDA and SVM, respectively. Both tables show results from DDSM and BSSA.

Table 3.4 presents the LDA classifier experimental results using structured micropatterns features ($\{f_0, f_1, f_{255}\}$, $\{f_0, f_3, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$ with 4 radii

Table 3.3: BSSA classification results from 10 runs of 5-fold cross validation over 5 balanced datasets (250 values), the highest performance of each feature set when considering all radius combinations using LDA. AUC is the mean AUC scores $\pm$ standard deviation.

Feature Set	Mean	p-value
$\{ f_0, f_1, f_{255} \}$	**0.885 $\pm$ 0.026**	-
$\{ f_0, f_3, f_{255} \}$	0.874 $\pm$ 0.021	0.6780
$\{ f_0, f_7, f_{255} \}$	0.867 $\pm$ 0.020	0.0700
$\{ f_1, f_{31}, f_{255} \}$	0.858 $\pm$ 0.015	0.0263
$\{ f_1, f_3, f_{128} \}$	0.856 $\pm$ 0.024	0.0003

levels and their combinations for both datasets. Each row of the table corresponds to the classification results of radii levels 1, 2, 3, 4 and their combinations for each feature set. For a single radius level, $\{f_0, f_1, f_{255}\}$ obtained the highest AUC score of 0.957 (with radius 1) for DDSM and 0.814 (with radius 2) for BSSA. With two radii levels, radius 1 and radius 4 combination obtained highest score of 0.885 for BSSA. In case of DSSM, combining information from two radii levels did not result in increased the performance. With three radii levels, radius 1, radius 2 and radius 3 combination obtained highest score of 0.966 for DDSM and four radii level combination did not result in any increased classifier performance. In case of BSSA, three radii level combination and four radii level combination did not result in any increased classifier performance. These results are highlighted in bold in the table. Similar results for $\{f_0, f_3, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$) are also highlighted in bold in the table. Table 3.5 shows the experimental results for SVM-L classifier. Similar results were achieved with SVM-L classifier also. Figures 3.9, 3.10 and 3.11 show examples of 3D feature plot for SVM-linear for DDSM dataset1 for feature sets $\{f_0, f_1, f_{255}\}$, $\{f_0, f_3, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$ respectively.

3.8   Comparison with Related Publication

For a direct comparison of the proposed approach with the 'classic' histogram concatenation technique used in Llado et al. (2009) and Choi & Ro (2012), the mass ROI classification method proposed by Llado et al. was implemented and tested, on the selected DDSM subset of dense ROIs. There are two reasons for selecting the Llado et al. method over the Choi and Ro technique. First, Llado et al. used (like the proposed approach) a manual ROI selection while Choi and

Table 3.4: DDSM and BSSA mass classification using LDA classifier for different radius levels and their combinations for features $\{f_0, f_1, f_{255}\}$, $\{f_0, f_3, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$.

	DDSM - LDA								
	$\{f_0, f_1, f_{255}\}$			$\{f_0, f_3, f_{255}\}$			$\{f_0, f_7, f_{255}\}$		
Radius	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	**0.957**	**0.859**	**0.987**	**0.957**	**0.879**	**0.985**	**0.954**	**0.793**	**0.993**
2	0.903	0.778	0.940	0.911	0.800	0.923	0.900	0.785	0.945
3	0.882	0.735	0.917	0.908	0.793	0.903	0.891	0.773	0.903
4	0.861	0.734	0.854	0.880	0.762	0.849	0.860	0.742	0.852
1&2	0.955	0.843	0.979	**0.961**	**0.829**	**0.979**	0.951	0.840	0.971
1&3	0.955	0.889	0.968	0.955	0.872	0.968	0.955	0.868	0.975
1&4	0.956	0.900	0.970	**0.961**	**0.880**	**0.960**	**0.957**	**0.900**	**0.972**
2&3	0.903	0.811	0.913	0.902	0.808	0.913	0.902	0.811	0.912
2&4	0.912	0.829	0.919	0.904	0.818	0.912	0.909	0.824	0.910
3&4	0.917	0.822	0.895	0.905	0.829	0.880	0.906	0.842	0.878
1&2&3	**0.966**	**0.886**	**0.975**	0.958	0.878	0.974	0.947	0.871	0.965
1&2&4	0.959	0.885	0.978	0.958	0.888	0.965	0.953	0.881	0.961
1&3&4	0.952	0.886	0.978	0.957	0.883	0.965	0.951	0.879	0.974
2&3&4	0.906	0.806	0.916	0.901	0.828	0.900	0.907	0.810	0.915
1&2&3&4	0.960	0.873	0.976	0.960	0.882	0.962	0.944	0.864	0.969
	BSSA - LDA								
	$\{f_0, f_1, f_{255}\}$			$\{f_0, f_3, f_{255}\}$			$\{f_0, f_7, f_{255}\}$		
Radius	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.801	0.657	0.768	0.801	0.621	0.760	0.771	0.577	0.759
2	**0.814**	**0.627**	**0.792**	**0.810**	**0.635**	**0.775**	**0.817**	**0.679**	**0.773**
3	0.801	0.630	0.800	0.802	0.645	0.805	0.810	0.639	0.820
4	0.807	0.634	0.839	0.809	0.653	0.823	0.809	0.627	0.844
1&2	0.846	0.750	0.824	0.824	0.650	0.803	0.824	0.648	0.805
1&3	0.855	0.721	0.863	0.852	0.690	0.861	0.848	0.671	0.857
1&4	**0.885**	**0.748**	**0.860**	**0.874**	**0.722**	**0.874**	**0.867**	**0.705**	**0.860**
2&3	0.824	0.652	0.819	0.819	0.666	0.818	0.828	0.683	0.833
2&4	0.849	0.701	0.848	0.838	0.699	0.859	0.842	0.688	0.837
3&4	0.856	0.682	0.823	0.817	0.668	0.832	0.831	0.685	0.828
1&2&3	0.833	0.719	0.848	0.822	0.674	0.836	0.827	0.660	0.843
1&2&4	0.858	0.739	0.840	0.847	0.703	0.851	0.851	0.676	0.852
1&3&4	0.860	0.728	0.841	0.857	0.683	0.847	0.842	0.670	0.845
2&3&4	0.838	0.679	0.808	0.828	0.666	0.812	0.826	0.642	0.824
1&2&3&4	0.840	0.712	0.826	0.833	0.703	0.856	0.815	0.674	0.846

Ro used automatic ROI selection. Secondly, the performance of the Llado et al. method was better than the Choi and Ro approach. For the purpose of a fair comparison of results, implementation guidelines given in Llado et al. (2009) were strictly followed. Each ROI was divided into 5x5 squared local image patches. Then, uniform LBP features were extracted from these local regions with P=8 and R=1 and their LBP histograms were concatenated. As suggested by the

Table 3.5: DDSM and BSSA mass classification using SVM-L classifier for different radius levels and their combinations for features $\{f_0, f_1, f_{255}\}$, $\{f_0, f_3, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$.

	DDSM - SVM-L								
	$\{f_0, f_1, f_{255}\}$			$\{f_0, f_3, f_{255}\}$			$\{f_0, f_7, f_{255}\}$		
Radius	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	**0.953**	**0.889**	**0.975**	**0.958**	**0.907**	**0.965**	**0.949**	**0.870**	**0.978**
2	0.905	0.764	0.948	0.917	0.802	0.910	0.911	0.766	0.948
3	0.872	0.593	0.970	0.907	0.740	0.909	0.897	0.644	0.966
4	0.846	0.609	0.918	0.882	0.746	0.862	0.858	0.667	0.906
1&2	0.952	0.881	0.967	0.957	0.889	0.958	0.953	0.884	0.956
1&3	0.960	0.890	0.954	0.957	0.896	0.957	0.962	0.903	0.957
1&4	**0.961**	**0.912**	**0.956**	**0.959**	**0.901**	**0.953**	**0.965**	**0.907**	**0.952**
2&3	0.918	0.812	0.925	0.906	0.813	0.909	0.918	0.832	0.930
2&4	0.920	0.841	0.909	0.910	0.820	0.896	0.919	0.852	0.907
3&4	0.900	0.763	0.910	0.904	0.786	0.893	0.907	0.802	0.902
1&2&3	**0.962**	**0.881**	**0.958**	0.958	0.892	0.963	0.958	0.893	0.954
1&2&4	0.961	0.906	0.968	0.957	0.900	0.960	0.964	0.905	0.957
1&3&4	0.959	0.900	0.957	0.956	0.894	0.957	0.962	0.907	0.957
2&3&4	0.917	0.845	0.905	0.907	0.814	0.902	0.921	0.860	0.916
1&2&3&4	0.961	0.901	0.965	0.957	0.895	0.963	0.961	0.901	0.957
	BSSA - SVM-L								
	$\{f_0, f_1, f_{255}\}$			$\{f_0, f_3, f_{255}\}$			$\{f_0, f_7, f_{255}\}$		
Radius	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.802	0.644	0.751	0.800	0.608	0.748	0.799	0.587	0.781
2	**0.817**	**0.616**	**0.796**	**0.811**	**0.621**	**0.778**	0.815	0.637	0.790
3	0.804	0.606	0.793	0.799	0.600	0.792	**0.818**	**0.638**	**0.805**
4	0.816	0.567	0.843	0.807	0.570	0.843	0.817	0.590	0.851
1&2	0.819	0.654	0.786	0.801	0.586	0.774	0.815	0.616	0.787
1&3	0.840	0.656	0.825	0.819	0.620	0.812	0.826	0.620	0.828
1&4	**0.855**	**0.672**	**0.852**	**0.848**	**0.646**	**0.853**	**0.852**	**0.622**	**0.854**
2&3	0.823	0.618	0.822	0.813	0.642	0.809	0.833	0.655	0.808
2&4	0.835	0.637	0.856	0.829	0.672	0.838	0.845	0.653	0.841
3&4	0.825	0.633	0.831	0.810	0.614	0.845	0.837	0.645	0.834
1&2&3	0.837	0.657	0.825	0.825	0.632	0.810	0.823	0.606	0.819
1&2&4	**0.857**	**0.677**	**0.858**	0.838	0.658	0.846	0.833	0.614	0.849
1&3&4	0.855	0.671	0.863	0.841	0.661	0.851	0.844	0.629	0.860
2&3&4	0.836	0.642	0.857	0.830	0.667	0.858	0.838	0.654	0.856
1&2&3&4	0.853	0.680	0.865	0.841	0.647	0.854	0.835	0.621	0.849

authors, this concatenated histogram was combined with the LBP histogram extracted from central 3×3 local regions to form the final texture descriptor, which resulted in 2006 features. Table 3.6 shows the comparison outcomes. It is clear that the proposed approach was significantly better with respect to feature dimensionality with comparable AUC score.
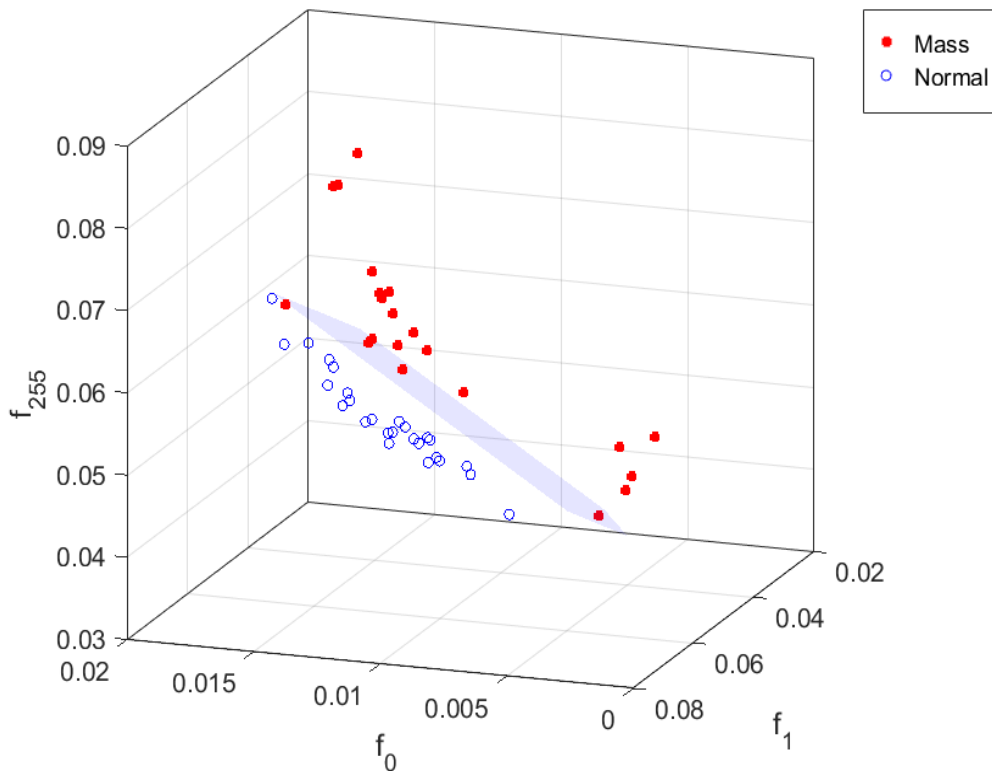
(a)

Figure 3.9: Example showing 3D feature plot for SVM-linear for $\{f_0, f_1, f_{255}\}$ for radius 1 for classifying mass and normal class in dataset1.

Table 3.6: AUC score comparison of the proposed approach with the approach introduced by Llado et al. (2009). Best results are highlighted in bold. DDSM, LDA result 0.957 and SVM-L result 0.965 obtained for $\{f_0, f_7, f_{255}\}$ at radius 1 and 4, BSSA, LDA result 0.885 and SVM-L result 0.855 obtained for $\{f_0, f_1, f_{255}\}$ at radius 1 and 4

| Approach | No. of Features | DDSM | | BSSA | |
		LDA	SVM-L	LDA	SVM-L
Llado et al. (2009)	2006	0.949	**0.972**	**0.910**	**0.918**
Proposed	**6**	**0.957**	0.965	0.885	0.855

3.9 Discussion

In this study, LBP was computed on individual pixel gray values to describe structured micropatterns. Mammographic images are noisy and pixel based LBP is susceptible to noise (Liu et al. 2016). Hence, a median filter was used to alleviate the noise influence on the classification procedure.
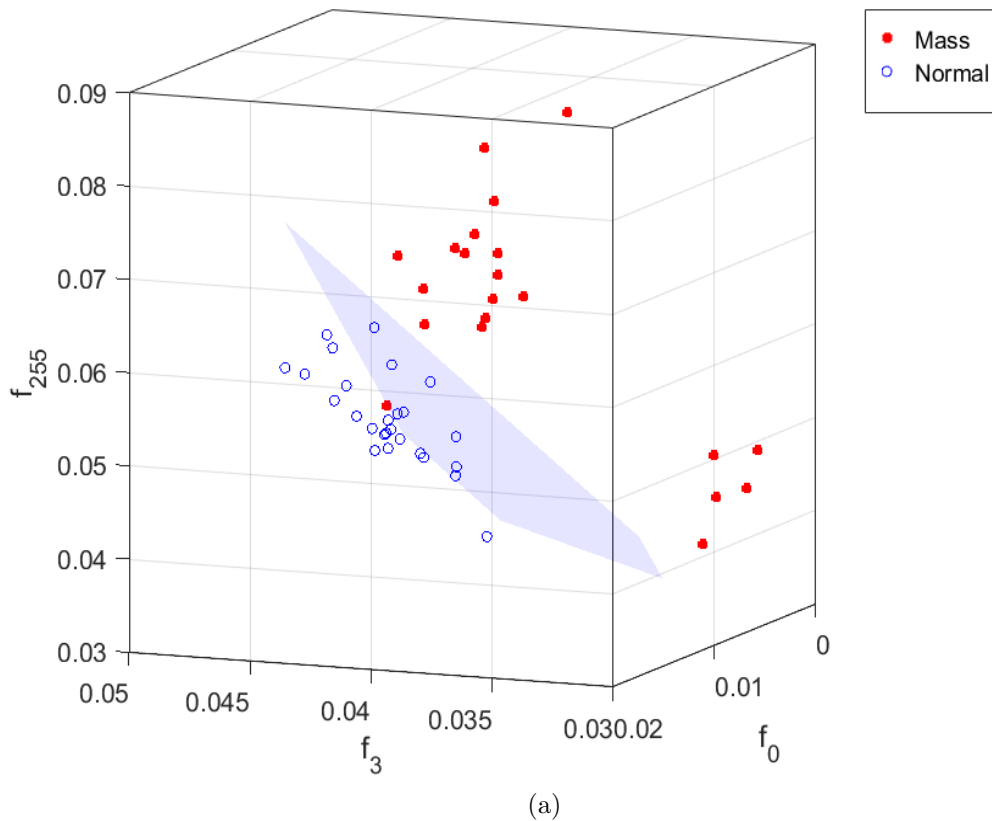
Figure 3.10: Example showing 3D feature plot for SVM-linear for $\{f_0, f_3, f_{255}\}$ for radius 1 for classifying mass and normal class in dataset1.

While analysing the best performing feature sets, it can be observed that the *spike* and *pit* combinations ($\{f_0, f_{255}\}$) with any other pattern from (f_1, f_3 f_7) gave the best results. Therefore, one could expect that this combination (*spike* and *pit*) has a high potential in discriminating a cancerous region from a normal region.

Our analysis shows that LBP feature dimensionality can be reduced significantly using efficient LBP features (see Table 3.4, 3.5), while retaining high texture classification accuracy. Recent studies have indicated that a comparable performance is achieved only with high feature dimensionality (240 to 6000 features) (Junior et al. 2009, Nascimento et al. 2013), carrying risk of classifier overfitting and lack of generalization. Feature dimensionality reduction is highly important, when considering scarcity of labeled data for dense breasts. To our knowledge, there is no study in the literature specifically focusing on dense breasts which may be due to lack of data set availability. We have visually analyzed more than 4000 mammograms across DDSM and BSSA to get 70 mass ROIs localized in dense background. This selection was evaluated by an experienced radiologist.
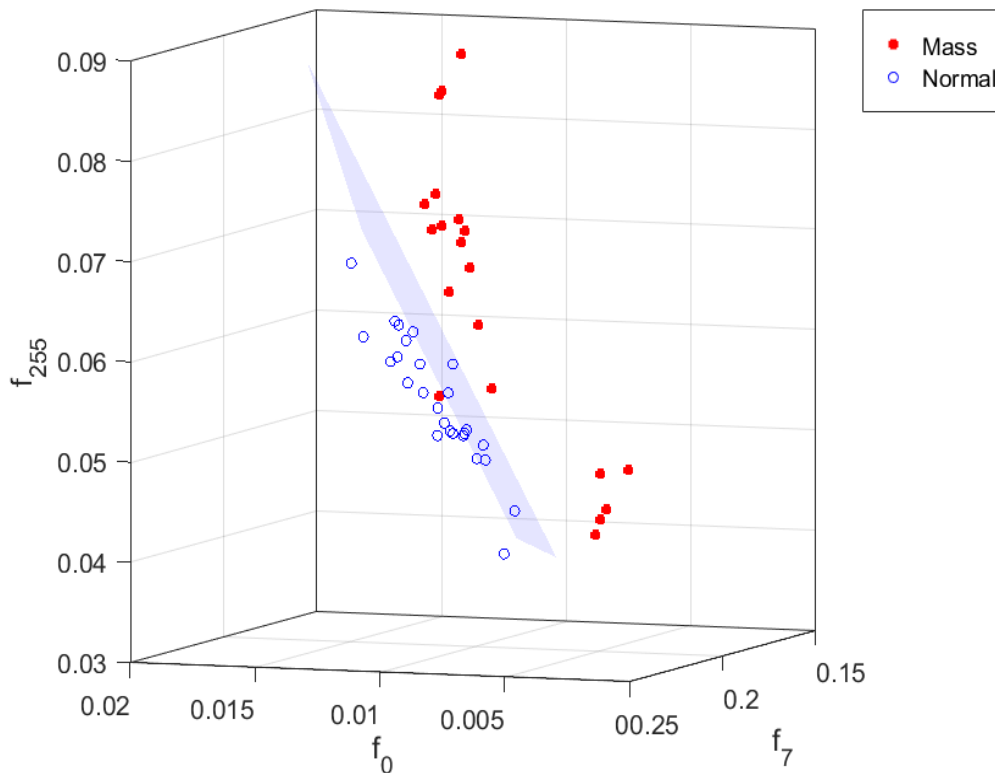
Figure 3.11: Example showing 3D feature plot for SVM-linear for $\{f_0, f_7, f_{255}\}$ for radius 1 for classifying mass and normal class in dataset1.

The efficient LBP features were validated with multiple classifiers (LDA and SVM) and achieved consistent results. Additionally, the Llado et al. (2009) approach was executed on the same ROIs as our approach to get a fair comparison and the outcome is shown in Table 3.6. The resultant performance was comparable with significant reduction in feature dimensionality (2006 features to 6 features). This is a significant achievement in terms of generalization. It is a well known fact from general theory of classification in pattern recognition that the capacity of a linear classifier is about twice the number of features. That is, the probability that a random distribution of classes will result in linear separability is very close to zero if the number of instances exceeds the number of features by more than twice.

In order to validate the performance of our method on unseen data, the test was conducted on a different database BSSA, which was not used in the feature selection procedure. The result shows that the classification performance on DDSM data is better than on BSSA data (0.965 for DDSM and 0.885 for BSSA). It may be partly because feature selection was done on a subset of the DDSM

database and this might have resulted in a slight bias. An AUC score of 0.885 for BSSA is a significantly good outcome for mass classification in dense breast using very few features.

However, the proposed approach has some limitations that need to be addressed. As the features are based on pixel based LBP, they are susceptible to noise and cannot represent macrostructures present in the image. Finally, these features do not reflect the spatial distribution and closeness of the patterns. To overcome these limitations, a new paradigm - the superpixel tessellation which can represent macrostructures in a image and is robust to noise - is presented in Chapter 4. Moreover, to reflect the spatial distribution and closeness of the patterns, a novel spatial connectivity graph model is developed in Chapter 5.

3.10 Conclusion

This chapter has presented a method for classification of mammographic dense regions as malignant masses or normal dense breast regions using structured micropatterns generated from LBP. The proposed method has been tested on 525 ROIs taken from two databases and compared with a recent state-of-the art mass classification outcome found in literature. Experimental results show that the proposed approach has a high potential to discriminate dense cancerous ROIs from normal dense breast ROIs. The typical histogram concatenation approach of extracting features from LBP was replaced by features generated from the efficient structured micropatterns. This resulted in a huge reduction of dimension of the feature space while keeping a high rate of mass classification. The efficient LBP features obtained the highest AUC score of 0.965 for DDSM and 0.885 for BSSA with only 6 features using LDA classifier. When compared with the recent state-of-the-art mass classification method, the proposed approach was significantly better in comparison to feature dimensionality and very comparable in terms of performance. However, as described in the discussion section, there are limitations for this approach such as susceptibility to noise, inability to represent macrostructures and their spatial distribution. These limitations are addressed in the next two chapters.

Chapter 4

Superpixel Texture Analysis

Mammographic images are noisy in nature and LBP computed on individual pixel grey values is sensitive to noise. In order to overcome these limitations, in this chapter we propose a simple and highly efficient approach, the superpixel-based LBP (SLBP), which can also capture image macrostructures. In the proposed approach, ROIs are transformed into superpixel tessellations. The superpixel tessellation represents the ROI in further texture analysis. The SLBP computed on the superpixels can better represent the macrostructures in the image. The nine rotational invariant uniform local binary patterns (structured micropatterns), described and analysed in Chapter 3, are defined on a superpixel grid structure in this chapter for analysing mass ROI classification. Similar to the previous chapter, class discrimination capability of the structured superpixel patterns is evaluated using AUC, sensitivity and specificity. Section 4.2 presents the proposed SLBP technique. Methodology is described in Section 4.3 and validity of the proposed method is evaluated using two datasets in Section 4.4. Finally, discussion and conclusions are provided in Section 4.5 and Section 4.6, respectively.

4.1   Dataset

The data used in this chapter for experiments are the same as described in Section 3.4. A total of 525 ROIs were used (301 extracted from DDSM and 224 from BSSA). All 525 ROIs were localized in dense backgrounds of breasts.

4.2 Superpixel Based Local Binary Patterns

Superpixel representation is an important preprocessing step in many computer vision algorithms. It is formed by locally grouping the pixels of an image and representing the group using a representative value. Superpixels simplify the image representation and make it more meaningful and easier to analyse. Superpixel representation has been used in various computer vision applications such as object recognition (Endres & Hoiem 2014), video analysis (Drucker & Mac-Cormick 2009), breast image segmentation (Chuand et al. 2015, Hao et al. 2012), and cell segmentation and tumor detection in histopathological images (Lee & Kim 2016, Ushizima et al. 2014). The main advantages of using superpixels are: better representation of the image by reducing redundancy, computational efficiency, robustness to noise (Achanta et al. 2012) and ability to describe texture macrostructures in the image (Liu et al. 2016). Superpixels can be generated by regular grid approach or by irregular grid approach.

In literature, few studies (Liao et al. 2007, Zhang et al. 2007, Liu et al. 2016) have used LBP based on regular grid superpixel approach for face recognition and scene image classification. Liao et al. (2007) and Zhang et al. (2007) used regular grid superpixel to compute local binary patterns for face recognition. The superpixel was represented by the mean value of the pixels grouped. Liu et al. (2016) proposed Median Robust Extended LBP (MRELBP), to compute LBP where the size of superpixels were different with different radii calculated. However, in all the above mentioned studies (Liao et al. 2007, Zhang et al. 2007, Hafiane et al. 2008, Liu et al. 2016), the feature vector was generated using traditional histogram (LBP label histogram) concatenation approach which does not recognize any specific structured patterns and hence leads to high dimensionality of the feature space (800 in Liu et al. (2016)). Another main disadvantage of the above studies using regular grid superpixel approach is, that the pixels in the group may not be homogeneous thus compromising the accuracy of macrostructure representation. In this study, we have used regular grid and irregular grid approach as described in Section 4.3.1 for generating superpixels. As mentioned previously, the nine rotational invariant uniform local binary patterns (structured micropatterns), described and analysed in Chapter 3, are defined on a superpixel grid structure in this chapter for analysing mass ROI classification. The histogram concatenation approach of extracting features from LBP utilised in (Liao et al. 2007, Zhang et al. 2007, Liu et al. 2016) is replaced in our study by features based on combinatorial properties of the structured superpixel patterns. This

allows for a huge reduction of the dimension of the feature space while keeping a high rate of mass classification. To our knowledge, this study is the first study of building LBP on irregular grid superpixel for mass classification.

4.2.1 The Proposed SLBP

In SLBP, superpixels take the role of pixels in LBP. Single pixel intensity values are replaced by the mean or median value of the superpixels and compared with the mean/median value of its k neighbouring superpixels. Given a centre superpixel s_c and the ϕ operator defined on superpixels (e.g., mean or median), the SLBP label is calculated as follows

$$SLBP(s_c) = \sum_{i=0}^{k-1} 2^i S(\phi(s_i) - \phi(s_c)), \tag{4.1}$$

where

$$S(\phi(s_i) - \phi(s_c)) = \begin{cases} 1, & \text{if } \phi(s_i) \geqslant \phi(s_c), \\ 0, & \text{if } \phi(s_i) < \phi(s_c). \end{cases} \tag{4.2}$$

The SLBP pattern is made rotation invariant by performing k bit-wise shifts and selecting the smallest value of the binary pattern as follows

$$SLBP(s_c^i) = min\{Rot(SLBP(s_c, i))|i = 0, 1, \ldots, k-1\}, \tag{4.3}$$

where $Rot(r, i)$ performs an i step circular bit-wise right shift on k-bit number r.

4.2.2 SLBP Neighbourhood Structure

Given superpixel s, the neighbourhood $N_s^{L,k}$ is defined for $L \in \{1, 2, 3\}$ and $k \in Z^+$. For each neighbourhood level L (we call it neighbourhood level instead of radius, because it is not fixed length in all directions), we find k neighbours by dividing the angular space $\theta = 360^0$ into k equal angles and drawing k radial lines from the centroid of each superpixel towards the boundary. For example, $k = 8$ yields the following quantization of $\theta : \{0, \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2, 7\pi/4\}$.

Figure 4.1 illustrates the proposed SLBP neighbourhood selection with central superpixel and neighbours marked. The first level neighbourhood was composed of neighbours that are the immediate neighbours of the centre superpixel at specified direction (θ) (see Figure 4.1 superpixel ⭐ and first level neighbours 🔴). In some cases, more than one radial line may pass through the same superpixel. For example, in Figure 4.1 two radial lines ($\theta = 7\pi/4$ and $\theta = 0$) pass through the same first level superpixel neighbour. The same is the case with radial lines passing through ($\theta = \pi$ and $\theta = 5\pi/4$). To find the second level neighbours, all first level neighbours are merged with the centre superpixel to form a bigger superpixel and superpixels connecting to that merged superpixel in particular directions were taken as the second level neighbours (see Figure 4.1 ➕). The same procedure was repeated for finding the third level neighbours (see Figure 4.1 ◀). By altering the values of k and L, SLBP operators for any spatial resolution and for any quantization of the angular space can be achieved. Observe that since superpixels can share neighbours in more than one direction, the mean/median value of the shared superpixel is used to represent all those directions. In addition, if any particular neighbourhood level does not exist (boundary conditions), then we ignore that superpixel from making the SLBP patterns at that particular level.
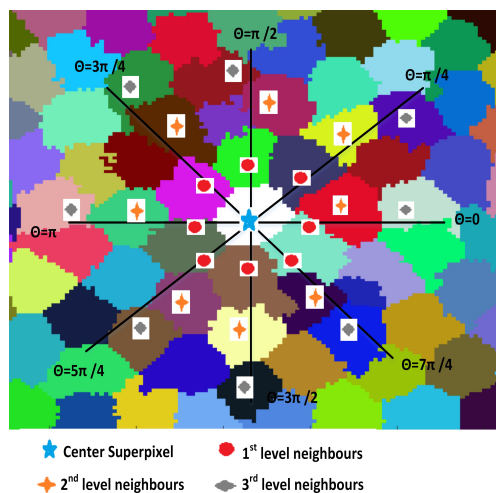
Figure 4.1: Example of SLBP neighbourhood showing $N_s^{1,8} \cup N_s^{2,8} \cup N_s^{3,8}$ where s is the center (in white) superpixel.

4.2.3 Structured Superpixel Patterns (*SpPatterns*)

Recall the nine uniform rotationally invariant local binary patterns: p_0 (00000000), p_1 (00000001), p_3 (00000011), p_7 (00000111), p_{15} (00001111), p_{31} (00011111), p_{63}

(00111111), p_{127} (01111111) and p_{255} (11111111) described in Section 3.3. In this chapter, these nine rotational invariant uniform local binary patterns are redefined as SLBP patterns on the superpixel grid structure to represent nine macrostructures as described in Sections 4.2.1 and 4.2.2. The dense mass ROI classification capability of these nine structured SLBP patterns are investigated in this chapter.

4.3   Methods

The proposed approach consists of the following steps: ROI extraction, superpixel generation, structured superpixel patterns extraction using the proposed SLBP technique, feature generation and classification using the features generated from structured superpixel patterns. Figure 4.2 shows the block diagram of the proposed approach. The data used in this chapter and ROI extraction mechanism, are the same as described in Section 3.4.
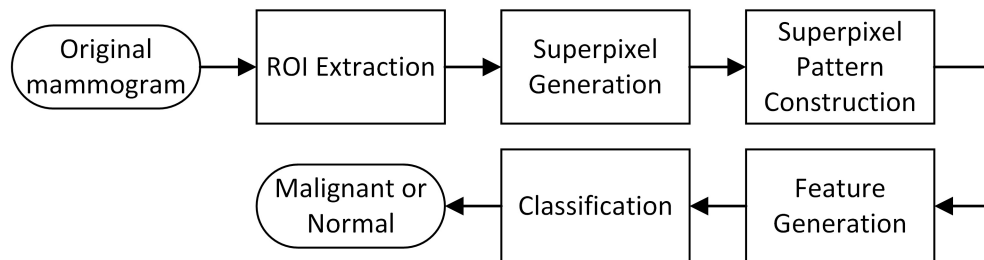
Figure 4.2: Block diagram of the proposed SLBP approach for mass classification.

4.3.1   Superpixel Generation

Three different superpixel generation algorithms were studied. First is a regular grid approach, where superpixels are generated by grouping the neighbouring pixels. In this approach, all the superpixels are of similar size and so intensity homogeneity within a superpixel cannot be guaranteed. Second is an irregular grid approach with controlled maximum size of superpixels, using the SLIC algorithm (Achanta et al. 2012). The superpixels generated using SLIC are partly homogenous and quasi similar in size. Third is an irregular grid approach using SRM algorithm (Nock & Nielsen 2004). The superpixels generated using SRM are highly homogeneous but their size approximation cannot be guaranteed.

Three different sizes of superpixels (9, 16 and 25 pixels) were generated. The regular grid superpixels of size 9, 16 and 25 were generated by grouping the 3×3,

4×4 and 5×5 neighbourhood of a center pixel respectively. SLIC superpixels were generated based on the initial grid structure (3×3, 4×4 and 5×5) and the parameter m that controls the relative weight between intensity and spatial proximity. The parameter m was set to the superpixel size to give equal weight to both intensity and spatial proximity. For SRM, the Q value was selected automatically for each ROI to generate superpixels of approximately 9, 16 and 25 pixels. To automatically select the right Q value, each ROI was segmented using the SRM algorithm for a specified range (1000 to 20000 with step 200) of Q values. Then for each value of Q, the median value of the superpixels size was computed. The smallest values of Q resulting in median superpixel sizes 9, 16 and 25 were selected. Figure 4.3 shows an example of superpixel generation using regular grid approach, SLIC and SRM.

4.3.2 Feature Generation

To reduce the computational complexity, the number of neighbours k was selected to be 8 and the level $L \in \{1, 2, 3\}$. The number of occurrences of each of the nine structured superpixel patterns ($SpPatterns$) generated using the proposed SLBP technique per unit area were used as features. That is, for each ROI we define a feature as follows:

$$f_i = \frac{|SpPattern_i|}{|ROI|}, \tag{4.4}$$

where $|.|$ stands for the number of elements in the set (number of superpixels), $i = 0, 1, 3, 7, 15, 31, 63, 127, 255$ and $SpPattern_i$ is the ith pattern. Using notation introduced in Section 4.2.3 we have features f_0, f_1, f_3, f_7, f_{15}, f_{31}, f_{63}, f_{127}, f_{255} corresponding to the prominent nine $SpPatterns$. This results in nine features for each of the 3 levels. The features generated are independent of ROI size as it was normalized with the number of superpixels in each ROI.

4.3.3 Feature Selection

Similar to Chapter 3, exhaustive search was used for feature selection. The study was conducted with superpixel size $\approx 3 \times 3$ (approximate only for irregular grid), k = 8, L = 1 and superpixels generated using three different approaches (regular grid, SLIC and SRM) and two different representations for superpixel (mean and
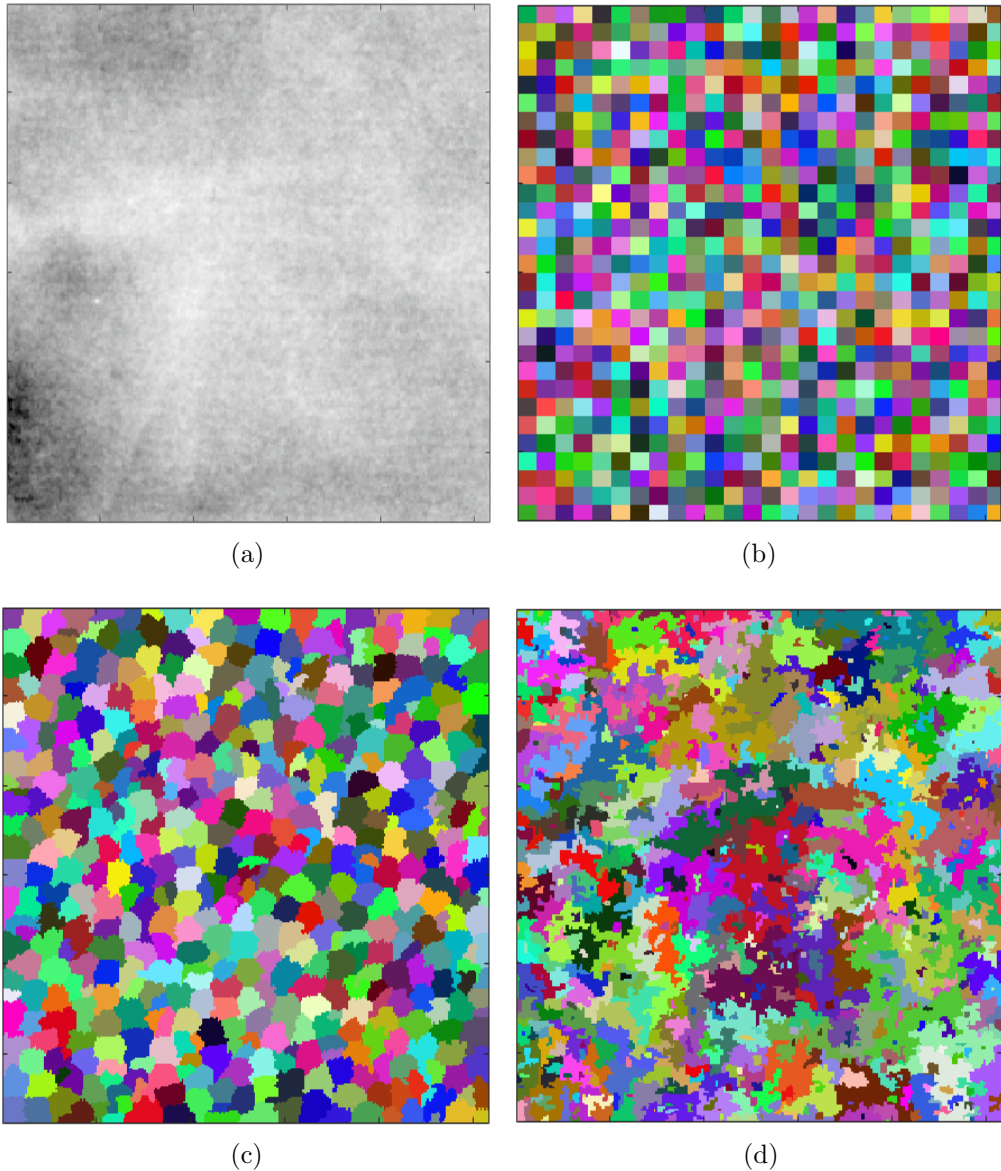
Figure 4.3: An example of superpixel generation (a) original (b) regular grid approach (c) SLIC and (d) SRM.

median representation). With nine features generated from structured SLBP patterns there are $2^9 - 1 = 511$ feature combinations. The dataset (DDSM dataset1 (93 ROIs)-see Section 3.4), learning algorithm (LDA), performance evaluation method (AUC) and validation method (ten independent runs of 5-fold cross validation) used here are also same as described in Section 3.5.3.

While analysing 511 feature combinations, LBP on regular grid superpixels (Regular-SLBP) with mean representation, the $\{f_1, f_7, f_{15}, f_{127}\}$ feature combination obtained the highest AUC of 0.685; and with median representation, the

$\{f_0, f_7, f_{15}, f_{255}\}$ combination obtained the highest AUC of 0.930. In the case of SLIC-SLBP with mean representation, the $\{f_0, f_1, f_{15}, f_{127}, f_{255}\}$ combination obtained the highest AUC of 0.794; and with median representation, the $\{f_3, f_{31}, f_{63}, f_{127}\}$ obtained the highest AUC of 0.825. For SRM-SLBP with mean representation, the $\{f_0, f_1, f_{31}, f_{63}, f_{255}\}$ combination obtained the highest AUC of 0.938; and with median representation, the $\{f_0, f_7, f_{63}, f_{255}\}$ combination obtained the highest AUC of 0.941. Table 4.1 shows these results. The table also shows a comparison of highest scored mean and median features set with its corresponding median and mean feature set, for a fair comparison.

Table 4.1: Performance comparison for DDSM mass ROI (dataset1-93 ROIs) classification using regular grid, SLIC and SRM approaches with mean and median representation.

	Regular-SLBP		SLIC-SLBP		SRM-SLBP	
	Highest AUC Scored Feature Sets					
	Feature Set	AUC	Feature Set	AUC	Feature Set	AUC
Mean	$\{f_1, f_7, f_{15}, f_{127}\}$	0.685	$\{f_0, f_1, f_{15}, f_{127}, f_{255}\}$	0.794	$\{f_0, f_1, f_{31}, f_{63}, f_{255}\}$	**0.938**
Median	$\{f_0, f_7, f_{15}, f_{255}\}$	0.930	$\{f_3, f_{31}, f_{63}, f_{127}\}$	0.825	$\{f_0, f_7, f_{63}, f_{255}\}$	**0.941**
	Comparison					
	Feature Set	AUC	Feature Set	AUC	Feature Set	AUC
Mean	$\{f_1, f_7, f_{15}, f_{127}\}$	0.685	$\{f_0, f_1, f_{15}, f_{127}, f_{255}\}$	0.794	$\{f_0, f_1, f_{31}, f_{63}, f_{255}\}$	0.938
Median	$\{f_1, f_7, f_{15}, f_{127}\}$	0.724	$\{f_0, f_1, f_{15}, f_{127}, f_{255}\}$	0.767	$\{f_0, f_1, f_{31}, f_{63}, f_{255}\}$	0.932
Mean	$\{f_0, f_7, f_{15}, f_{255}\}$	0.600	$\{f_3, f_{31}, f_{63}, f_{127}\}$	0.695	$\{f_0, f_7, f_{63}, f_{255}\}$	0.892
Median	$\{f_0, f_7, f_{15}, f_{255}\}$	0.930	$\{f_3, f_{31}, f_{63}, f_{127}\}$	0.825	$\{f_0, f_7, f_{63}, f_{255}\}$	0.941

The experimental results show that SLBP computed on SRM superpixels can describe the mass characteristics better than the regular grid and SLIC. It can be observed that performance of the homogeneous superpixel based approach using SRM was comparable between mean and median representation while non-homogeneous superpixel (regular grid and SLIC) varies significantly between mean and median representation. Even though Regular-SLBP with superpixel size 3×3 and median representation was comparable with SRM-SLBP, Regular-SLBP performance degraded significantly for the superpixel sizes 4×4 (AUC = 0.659) and 5×5 (AUC = 0.712) with selected features, as the variability of the pixels in the superpixel increases with size. In case of SRM-SLBP, performance remains similar with superpixel sizes 4×4 (AUC = 0.930) and 5×5 (AUC = 0.917) with selected features. Based on this analysis, the SRM superpixel approach with median representation was selected as the best choice for mass classification in this chapter and the other two approaches were not selected for further analysis.

As mentioned, the feature set $\{f_0, f_7, f_{63}, f_{255}\}$ generated using SRM su-

perpixel approach with median representation achieved the highest score 0.941. Pairwise T-test was conducted to see whether this highest score was significantly different from others at a significance level of 0.05. All feature combinations of 1, 2, 3 and 4 features (in total 255 feature combinations) from SRM-SLBP with median representation were tested (one at a time) against the highest scored feature set $\{f_0, f_7, f_{63}, f_{255}\}$. From the T-test results, 24 feature combinations were found to be not statistically different in performance from the highest performing feature set $\{f_0, f_7, f_{63}, f_{255}\}$. Out of these 24 feature combinations, four feature sets were selected $\{f_0, f_7, f_{255}\}$, $\{f_1, f_7, f_{255}\}$, $\{f_1, f_{31}, f_{255}\}$ and $\{f_0, f_1, f_{255}\}$ for further analysis as their dimension was less than the highest scored feature set and not statistically different in performance.

4.4 Experimental Results

Similar to Chapter 3, the ROI classification power of the selected features was evaluated on the DDSM database (301 ROIs - all 5 datasets) and the unseen BSSA database (224 ROIs - all 5 datasets). Similarly ten independent runs of 5-fold cross validation were applied for each of the five data sets of BSSA and DDSM to ensure stable classification results. The reported results are averages of 10 runs over 5 balanced data sets. In addition to neighbourhood level 1, neighbourhood levels 2, 3 and their combinations were analysed.

The 4 three dimensional feature sets identified in Section 4.3.3 were applied to unseen BSSA data using LDA and their highest AUC scores (when considering all radius combinations) are shown in Table 4.2. Out of these 4 sets, the feature set $\{f_0, f_1, f_{255}\}$ achieved the highest AUC score of 0.924. As shown in the table, the feature set $\{f_0, f_7, f_{255}\}$ also performed at the same level. Hence we have presented the detailed results for these two feature sets $\{f_0, f_1, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$ for three superpixel sizes ($\approx(3 \times 3)$, $\approx(4 \times 4)$ and $\approx(5 \times 5)$), two datasets (DDSM and BSSA) and two classifiers (LDA and SVM).

Table 4.3 presents the experimental results using SLBP features ($\{f_0, f_1, f_{255}\}$) for DDSM and BSSA using LDA classifier for different superpixel sizes and neighbourhood levels. Each row of the table corresponds to the classification results of neighbourhood levels 1, 2, 3 and their combinations for feature set $\{f_0, f_1, f_{255}\}$. For example, each level (1, 2 and 3) has 3 features $\{f_0, f_1, f_{255}\}$ generated at respective levels. Combination of levels 1 and 2 has 6 total features ($\{f_0, f_1, f_{255}\}$ for level 1 and $\{f_0, f_1, f_{255}\}$ for level 2). Similarly combination

Table 4.2: BSSA classification results from 10 runs of 5-fold cross validation over 5 balanced datasets (250 values), the best performance of each feature set. Mean is the mean of AUC scores.

Feature Set	Mean	p-value
$\{ f_0, f_1, f_{255}\}$	0.924	-
$\{ f_0, f_7, f_{255}\}$	0.915	0.65
$\{ f_1, f_7, f_{255}\}$	0.869	7.12E-10
$\{ f_1, f_{31}, f_{255}\}$	0.851	1.50E-14

of levels 1 & 2 & 3 have nine total features. The highest AUC score of 0.961 for DDSM and AUC score of 0.924 for BSSA with six features (superpixel size 3×3, level 1 and 3 combination) were achieved. Similarly, Table 4.4 presents the experimental results for the SVM-L classifier for the same feature set. The highest scores are highlighted in bold in the table.

Table 4.5 and 4.6 show the experimental results using feature set $\{f_0, f_7, f_{255}\}$ for LDA and SVM-L classifiers respectively. The highest scores are highlighted in bold in the table and they are comparable with results from feature set $\{f_0, f_1, f_{255}\}$. Figure 4.4 shows the ROC curves for LDA and SVM-L classifiers for DDSM and BSSA datasets.

4.5 Discussion

In this chapter, LBP was computed on superpixels to describe the structured superpixel patterns. The structured superpixel patterns capture the local region appearance, spatial structure and their density forms the feature descriptor. The superpixels were generated using regular and irregular grid approaches and represented using both mean and median values of the pixels grouped in the superpixels.

Superpixel sizes starting from 3×3 were considered, as sizes below that (2×2) are expected to be susceptible to noise. It was also found that performance degraded with larger superpixel size, and so superpixel sizes beyond 5×5 were not considered. For Regular-SLBP, median represents superpixels more effectively than mean (see Table 4.1). This result is expected as regular grid superpixels consist of non homogeneous pixels and median is a better representation than mean, since it is more robust to outliers (noisy pixels). In the case of Irregular-

Table 4.3: Comparison of AUC scores obtained for selected features $\{f_0, f_1, f_{255}\}$ for DDSM and BSSA mass classification using LDA classifier for different superpixel sizes and neighbourhood levels).

	SRM-SLBP $\{f_0, f_1, f_{255}\}$								
	DDSM								
	$\approx (3 \times 3)$			$\approx (4 \times 4)$			$\approx (5 \times 5)$		
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.937	0.821	0.909	0.930	0.817	0.896	0.916	0.785	0.897
2	0.853	0.742	0.861	0.800	0.697	0.845	0.759	0.607	0.803
3	0.712	0.590	0.762	0.639	0.434	0.790	0.546	0.293	0.808
1&2	0.923	0.851	0.890	0.932	0.826	0.902	0.948	0.853	0.925
1&3	**0.961**	**0.852**	**0.919**	0.944	0.851	0.908	0.950	0.823	0.918
2&3	0.907	0.842	0.866	0.887	0.774	0.848	0.809	0.663	0.795
1&2&3	0.946	0.855	0.905	0.942	0.850	0.922	0.946	0.831	0.920
	BSSA								
	$\approx (3 \times 3)$			$\approx (4 \times 4)$			$\approx (5 \times 5)$		
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.839	0.759	0.769	0.779	0.670	0.776	0.761	0.631	0.770
2	0.770	0.612	0.793	0.808	0.606	0.817	0.747	0.533	0.794
3	0.848	0.629	0.821	0.827	0.604	0.833	0.881	0.673	0.873
1&2	0.874	0.777	0.796	0.870	0.710	0.824	0.795	0.680	0.800
1&3	**0.924**	**0.812**	**0.878**	0.858	0.688	0.840	0.900	0.730	0.889
2&3	0.871	0.671	0.868	0.840	0.646	0.817	0.883	0.699	0.893
1&2&3	0.906	0.777	0.859	0.875	0.729	0.845	0.870	0.699	0.883

SLBP both measures give similar scores. Even though Regular-SLBP with superpixel size 3×3 and median representation was comparable with SRM-SLBP, Regular-SLBP performance degraded significantly for the superpixel sizes 4×4 (AUC $= 0.659$) and 5×5 (AUC $= 0.712$) with selected features, as the variability of the pixels in the superpixel increases with size. Hence Regular-SLBP was not considered for further analysis.

The experimental results shows that similar to micro-structures (Chapter 3), macro-structures generated from homogeneous superpixel representation are successful in identifying the signs of cancer in dense mammograms. With only 6 features f_0, f_1, f_{255} at level 1 and 3 (Table 4.3), an AUC score of 0.961 on DDSM and 0.924 on unseen BSSA data was achieved with LDA classifier and similar performance with SVM-L classifier (see Table 4.4). Overall it is clear from the experimental results that structured superpixel patterns generated from homogeneous SRM superpixel representation has a high potential in classifying dense mass ROIs.

While analysing the best performing feature sets ($\{f_0, f_1, f_{255}\}$ and $\{f_0, f_7, f_{255}\}$), it can be observed that the *spike* and *pit* superpixel combination ($\{f_0,$

Table 4.4: Comparison of AUC scores obtained for selected features $\{f_0,\ f_1,\ f_{255}\}$ for DDSM and BSSA mass classification using SVM-L classifier for different superpixel sizes and neighbourhood levels.

SRM-SLBP $\{f_0,\ f_1,\ f_{255}\}$								
DDSM								
	$\approx(3\times3)$			$\approx(4\times4)$			$\approx(5\times5)$	
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.939	0.850	0.907	0.930	0.814	0.889	0.917	0.760	0.897
2	0.858	0.750	0.858	0.802	0.682	0.855	0.763	0.564	0.834
3	0.717	0.582	0.787	0.633	0.422	0.792	0.526	0.184	0.857
1&2	0.924	0.828	0.908	0.930	0.817	0.905	0.945	0.870	0.893
1&3	**0.947**	**0.848**	**0.905**	0.947	0.841	0.905	0.947	0.828	0.910
2&3	0.894	0.789	0.874	0.880	0.758	0.867	0.811	0.637	0.817
1&2&3	0.948	0.849	0.913	0.948	0.864	0.924	0.953	0.858	0.906
BSSA								
	$\approx(3\times3)$			$\approx(4\times4)$			$\approx(5\times5)$	
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.844	0.723	0.786	0.784	0.668	0.775	0.765	0.637	0.773
2	0.786	0.643	0.777	0.807	0.586	0.844	0.751	0.539	0.814
3	0.866	0.595	0.850	0.839	0.589	0.839	0.887	0.673	0.873
1&2	0.863	0.739	0.811	0.849	0.685	0.819	0.801	0.675	0.790
1&3	**0.916**	**0.795**	**0.845**	0.867	0.698	0.848	0.906	0.753	0.881
2&3	0.882	0.643	0.873	0.852	0.634	0.839	0.896	0.741	0.887
1&2&3	0.911	0.770	0.849	0.872	0.699	0.844	0.886	0.730	0.868

$f_{255}\}$) with one other pattern (f_1 or f_7) gave the best results. Similar results were observed in the previous chapter (Chapter 3). This shows that these combinations of structural patterns have high potential in discriminating cancerous regions from normal regions in micro (pixel) level and macro (superpixel) level representations.

The proposed approach is based on superpixel, which is inherently robust to noise and an explicit preprocessing step for denoising was not required. The absence of denoising is a significant advantage because local spatial information is important for texture recognition and denoising potentially suppresses some important local texture information. The performance comparison between the pixel-based features (Chapter 3) and the superpixel based features also shows the same. The superpixel based one achieved an AUC score of 0.924 on unseen BSSA data (see Table 4.3) while pixel one achieved only an AUC score of 0.885 (see Table 3.4).

The proposed approach has another advantage of no parameter tuning. In this study, only the SRM technique had a parameter Q which was selected automatically for each image.

The superpixel tessellation, introduced in this chapter, can lead to a range

Table 4.5: Comparison of AUC scores obtained for selected features $\{f_0,\ f_7,\ f_{255}\}$ for DDSM and BSSA mass classification using LDA classifier for different superpixel sizes and neighbourhood levels .

SRM-SLBP $\{f_0,\ f_7,\ f_{255}\}$								
DDSM								
	$\approx(3\times3)$			$\approx(4\times4)$			$\approx(5\times5)$	
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.921	0.812	0.909	0.919	0.817	0.899	0.908	0.784	0.880
2	0.833	0.716	0.823	0.727	0.578	0.780	0.692	0.505	0.783
3	0.583	0.360	0.788	0.490	0.198	0.818	0.461	0.190	0.819
1&2	0.912	0.821	0.905	0.915	0.815	0.911	0.946	0.841	0.927
1&3	**0.951**	**0.840**	**0.904**	0.949	0.865	0.912	0.942	0.794	0.901
2&3	0.887	0.786	0.857	0.837	0.734	0.785	0.750	0.564	0.795
1&2&3	0.943	0.831	0.918	0.945	0.866	0.923	0.944	0.816	0.917
BSSA								
	$\approx(3\times3)$			$\approx(4\times4)$			$\approx(5\times5)$	
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.824	0.735	0.759	0.778	0.662	0.773	0.760	0.648	0.779
2	0.784	0.615	0.797	0.813	0.602	0.827	0.779	0.571	0.829
3	0.860	0.659	0.838	0.833	0.631	0.815	0.882	0.707	0.861
1&2	0.859	0.771	0.802	0.867	0.746	0.828	0.789	0.656	0.817
1&3	**0.915**	**0.782**	**0.871**	0.853	0.708	0.839	0.884	0.737	0.878
2&3	0.867	0.667	0.857	0.815	0.621	0.834	0.866	0.674	0.875
1&2&3	0.898	0.763	0.874	0.879	0.753	0.854	0.867	0.719	0.874

of effective texture descriptors with high potential for texture classification. One such method is proposed in this chapter and another one in next chapter, there could be much more that is yet to be explored. Finally, as features explored in this chapter are based on density of structured superpixel patterns, it solves the limitations of the previous chapter (Chapter 3) such as, susceptibility to noise and inability to represent macrostructures in the image. However, it still lacks addressing the spatial distribution and closeness of the patterns, which is solved in the next chapter (Chapter 5).

4.6 Conclusion

In this study, a novel solution to extract texture features for characterization of mammographic dense masses was proposed. Structured superpixel patterns were computed on regular and irregular grid superpixels using SLBP technique. The proposed method has been tested on two databases. Experimental results show that the structured superpixel patterns computed from homogeneous superpixels have a high potential to discriminate dense cancerous ROIs from normal dense

Table 4.6: Comparison of AUC scores obtained for selected features $\{f_0, f_7, f_{255}\}$ for DDSM and BSSA mass classification using SVM-L classifier for different superpixel sizes and neighbourhood levels .

SRM-SLBP $\{f_0, f_7, f_{255}\}$								
DDSM								
$\approx(3 \times 3)$			$\approx(4 \times 4)$			$\approx(5 \times 5)$		
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.917	0.801	0.916	0.919	0.786	0.914	0.910	0.781	0.885
2	0.827	0.688	0.834	0.713	0.508	0.823	0.672	0.432	0.795
3	0.543	0.280	0.817	0.450	0.082	0.920	0.447	0.065	0.929
1&2	0.906	0.799	0.913	0.919	0.782	0.925	0.942	0.832	0.913
1&3	0.940	0.810	0.913	**0.950**	**0.848**	**0.935**	0.945	0.792	0.907
2&3	0.876	0.713	0.882	0.835	0.715	0.777	0.731	0.525	0.789
1&2&3	0.929	0.809	0.916	0.954	0.883	0.943	0.953	0.856	0.908
BSSA								
$\approx(3 \times 3)$			$\approx(4 \times 4)$			$\approx(5 \times 5)$		
Levels	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.824	0.682	0.771	0.788	0.654	0.777	0.765	0.646	0.774
2	0.800	0.643	0.787	0.817	0.588	0.846	0.790	0.584	0.833
3	0.861	0.595	0.852	0.838	0.639	0.818	0.881	0.699	0.856
1&2	0.852	0.739	0.797	0.854	0.694	0.834	0.795	0.670	0.819
1&3	**0.909**	**0.749**	**0.857**	0.860	0.699	0.857	0.878	0.747	0.866
2&3	0.885	0.683	0.861	0.845	0.650	0.855	0.878	0.692	0.885
1&2&3	0.899	0.749	0.848	0.881	0.737	0.852	0.874	0.748	0.863

breast ROIs. The structured SLBP features f_0, f_1 and f_{255} with level 1 and 3 information yielded the highest AUC score of 0.961 for DDSM and 0.924 for BSSA using LDA classifier. The proposed approach has the advantages of high discriminative power, noise robustness, no parameter tuning and better texture representation.
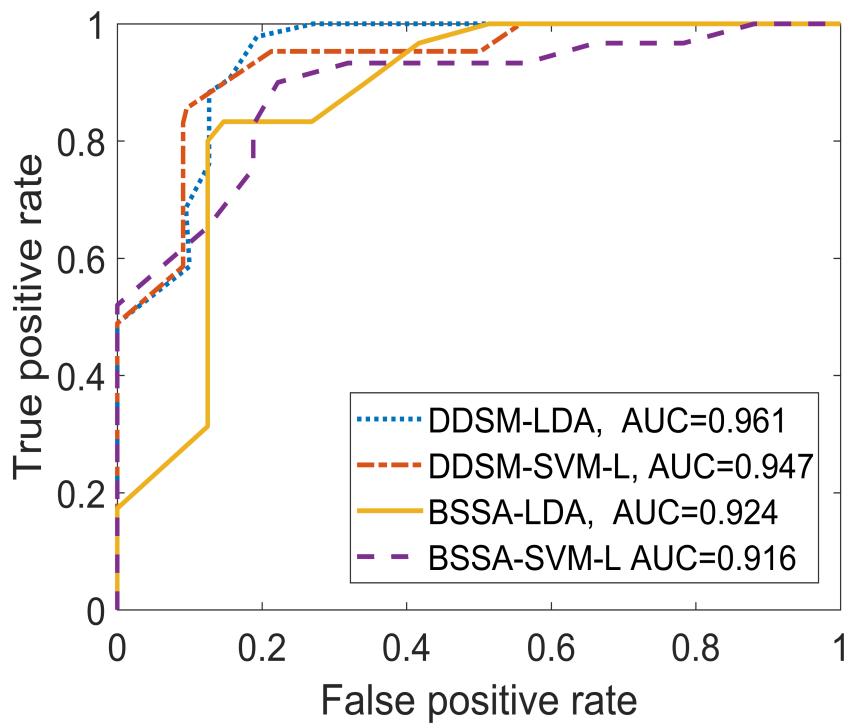
Figure 4.4: ROC curves illustrating the effectiveness of selected features for ROI classification for DDSM and BSSA datasets. The feature set $\{f_0, f_1, f_{255}\}$ was used to compute the ROC curves.

Chapter 5

Topological Modeling of Superpixel Patterns

In the previous two chapters, structured micro (pixel) and macro (superpixel) patterns, were described and their potential shown in discriminating dense mammographic ROIs. The proportion/density of the structured micropatterns and macro-patterns in each ROI were taken as the feature for classification. In this chapter, a novel method is proposed based on the topology of structured superpixel patterns. Graph models are constructed on structured superpixel patterns using morphological dilation to represent the spatial connectivity relationship between the structured superpixel patterns within a ROI. A set of topological features are generated from the proposed graph models and ROI classification is performed based on these features. Section 5.2 provides a brief review of various graph based approaches for image analysis. Section 5.3 describes the methodology adopted in this study. Section 5.4 details the experimental results obtained followed by comparison with other recent state-of-the-art mass classification techniques in Section 5.5. Discussion and conclusion is provided in Section 5.6.

5.1 Dataset

The data used in this chapter for experiments are the same as described in Section 3.4. A total of 525 ROIs were used (301 extracted from DDSM and 224 from BSSA). All 525 ROIs were localized in dense backgrounds of breasts.

5.2 Graph based Approaches

In literature, numerous graph based approaches can be seen in histopathological image analysis where the spatial connectivity relationship between the cells is measured to distinguish between cancerous and healthy tissues (Gunduz et al. 2004, Oztan et al. 2013, Gunduz et al. 2004, Bilgin et al. 2010, 2007). Segmentation and detection of cell nucleus is the initial step for most of the cell graph based studies. The centroids of the cell nuclei forms the graph nodes and edges are formed based on the spatial proximity between the nuclei. Different topological properties are captured based on the cell graph features and classification of tissues is performed based on these graph features. Early studies on cell graphs were based on graph techniques like Voronoi tessellation and Delaunay triangulation (Stephen et al. 2000, Barbara et al. 1999). Barbara et al. (1999) constructed a Voronoi diagram on the image and then its Gabriel's Graph (GG) and Minimum Spanning Tree (MST) were built. The graph features were extracted from the Voronoi diagram, GG and MST. In (Stephen et al. 2000), Delaunay triangulation was built on the image and graph features were extracted from the Delaunay triangulation graph. Both studies performed cell nuclei segmentation and identification before graph construction. However, these conventional graph techniques (Voronoi and Delaunay) construct global graph which connects every node. Hence, from these graphs, only global information like edge length statistics can be extracted. Later, cell graph approaches were generalized by introducing/allowing arbitrary edges based on the pairwise distance relationship between the nodes (Oztan et al. 2013, Gunduz et al. 2004, Bilgin et al. 2010, 2007). Cells or cell nuclei form the graph nodes and an edge between a pair of nodes exists based on the spatial proximity/distance between them. Both local and global graph features can be extracted from these approaches.

In the field of mammography, graph methods are used for pectoral muscle identification (Ma et al. 2007), breast segmentation (Don et al. 2011, Susukida et al. 2008) and temporal analysis of mammograms (Ma et al. 2008). Recently, Chen et al. (2015) used spatial connectivity graphs (similar to cell graph approaches) for classification of microcalcifications. However, graph based approaches for mammographic mass texture analysis have not been explored in the literature. One of the reasons for that could be graph models on pixel based texture features are complex and computationally expensive.

As mentioned previously, female breast is made of different types of parenchymal patterns. Organization of these patterns are not random but are linked with

the underlying function or functional state. The presence of cancer/masses can change the normal organization of the tissue patterns. Hence, understanding the tissue patterns/structure organization and their relationships can be used to predict malfunctioning when the patterns start changing. From these observations, it can be hypothesized that the spatial relationships between the structured superpixel patterns found in mass regions may differ from those found in normal dense regions. Hence, graph theoretical features extracted from the superpixel pattern graph that reflect their topological properties could be used to distinguish mass regions from normal dense regions. This study was motivated by the above hypothesis and related literature (Chen et al. 2015, Oztan et al. 2013, Gunduz et al. 2004, Bilgin et al. 2010, 2007).

In the proposed approach, first, superpixel ROI tessellations are produced by the SRM technique (Nock & Nielsen 2004). Then, using SLBP described in the previous chapter (Section 4.2), nine structured superpixel patterns are generated from superpixel tessellations. Subsequently, superpixel patterns graphs are constructed based on these nine patterns. Finally, the topology/connectivity of these nine superpixel pattern graphs is analysed and used to extract feature vectors for discriminating cancer and healthy dense ROIs. Similar to the previous two chapters, class discrimination capability of features generated from the structured superpixel patterns graphs are evaluated using AUC, sensitivity and specificity.

5.3 Methodology

The proposed approach consists of the following steps: ROI extraction, SRM superpixel generation, superpixel patterns extraction, morphological operation, superpixel pattern graph generation, topological feature extraction and classification. Figure 5.1 shows the block diagram of the proposed approach. SRM superpixel generation ($\approx$(3 × 3)) and superpixel patterns extraction were the same as the previous chapter and hence are not detailed in this chapter.

5.3.1 Morphological Operation

Mathematical morphology is based on set theory and is used widely for the analysis of spatial structures from an image (Sonka et al. 2007). It uses a collection of operations to extract relevant structures of the image. This is achieved by
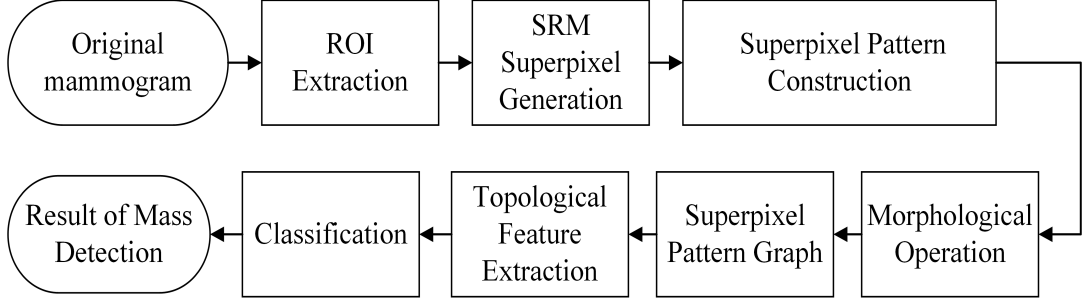
Figure 5.1: Block diagram of the proposed superpixel patterns topological modelling approach for mass classification.

probing the image with a structuring element (a small set). The shape of the structuring element can be decided based on prior knowledge of the shape of the object that needs to be processed (Chen 2013). Erosion and dilation are the two main morphological operations and many others are derived from these two operations. This study uses morphological dilation operation for constructing spatial connectivity graphs. Dilation operation increases an object boundary and is defined as follows. Assume X is a set and E is the structuring element, then dilation $(X \oplus E)$ is defined as:

$$(X \oplus E) = \bigcup_{e \in E} X_e, \tag{5.1}$$

where X_e denotes the translations of X by the vectors $-e$ of E. The dilation operation works on binary image. The binary images for structured superpixel patterns are defined as follows. Let p be one of the nine structured superpixel patterns defined in Section 4.2.3 and I be the image corresponding to one of the ROI. Define the binary image associated with the pattern p as follows

$$I_p(x, y) = \begin{cases} 1, & \text{if } s_{(x,y)} \in p, \\ 0, & otherwise, \end{cases} \tag{5.2}$$

where $s_{(x,y)}$ is the unique superpixel containing (x, y). Figure 5.2 shows an example of mass ROI and binary images corresponding to the nine structured superpixel patterns. Only level 1 neighbors are used to create the structured superpixel patterns. Visually different pattern clusters are clearly visible. The next section describes a technique used to analyze the characteristics of these clusters.

(a) Mass ROI (b) p_0 (c) p_1 (d) p_3

(e) p_7 (f) p_{15} (g) p_{31} (h) p_{63}

(i) P_{127} (j) p_{255}
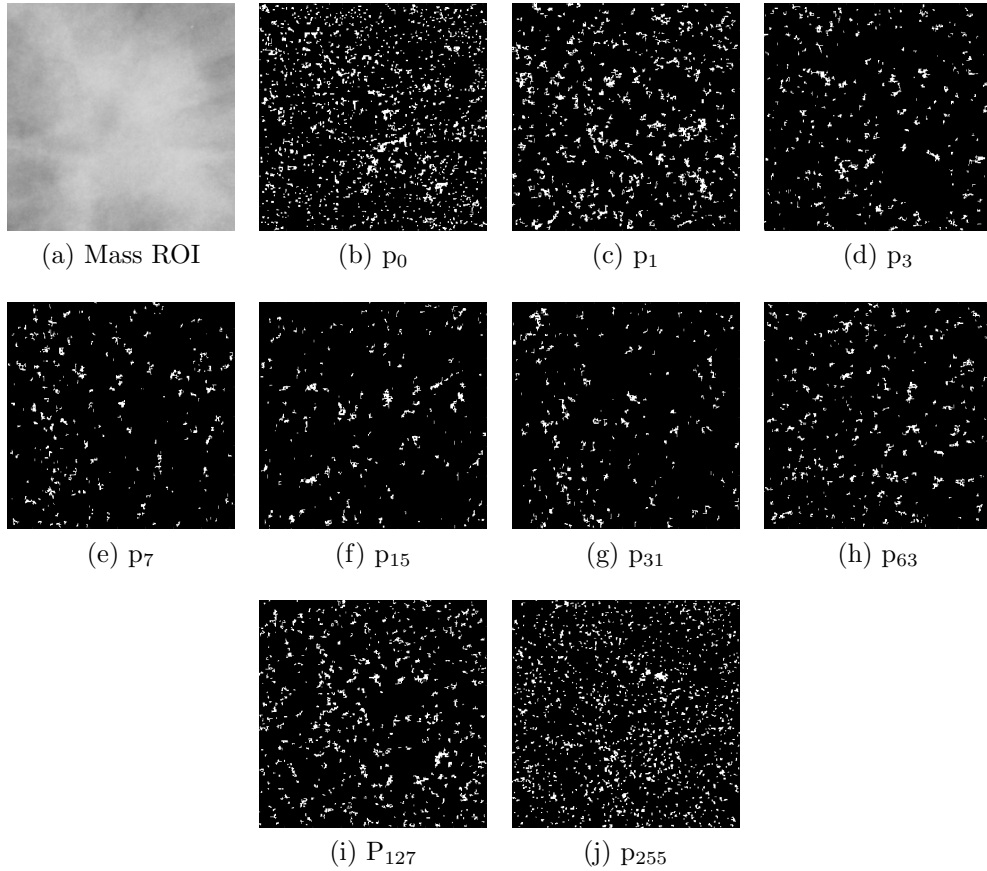
Figure 5.2: An example showing pattern-based binary images for an ROI. (Only level 1 neighbors were used.)

5.3.2 Superpixel Pattern Graph

To describe the texture of ROIs, a pattern graph for each of the binary images associated with the pattern (defined in Eq. 5.2) was built. The graph was generated utilizing spatial connectivity between patterns within a cluster. The connectivity was established by means of morphological dilation (with a disk shaped structural element) applied to individual superpixels at two scales: scale 1 - using one superpixel size as radius (9 pixels), and scale 2 - using two superpixel size as radius (18 pixels). In the graph, each node represents the structured superpixel, and an edge is added between two nodes if the two corresponding superpixels are connected or overlap with each other in the binary image plane. The adjacency matrix was used to encode the superpixel pattern graph. Let M be the adjacency matrix, then $M_{xy} \in \{0, 1\}, x, y = 1, 2, \ldots k$, where k is the number of nodes in graph and $M_{xy} = 1$ if two nodes i and j are connected, otherwise $M_{xy} = 0$. Figure 5.3 and 5.4 show an example of a mass ROI and its superpixel pattern graph constructed at scale 1 and 2 respectively.

(a) Mass ROI   (b) p_0 (*spike*)   (c) p_{255} (*pit*)
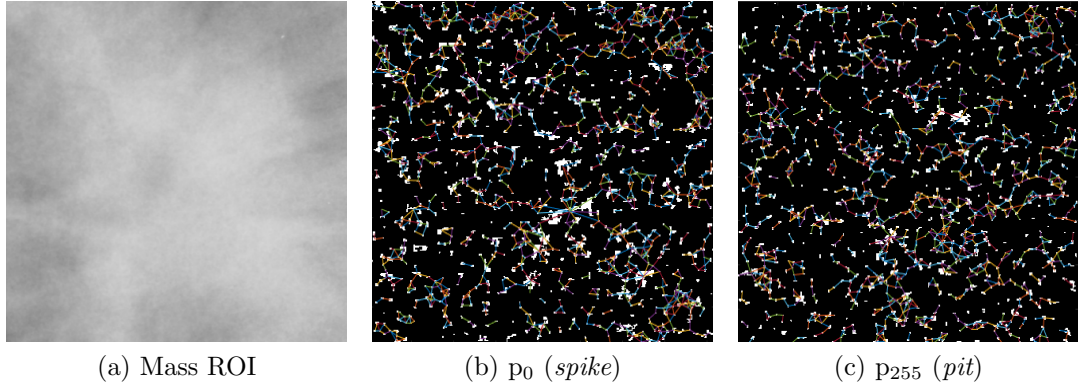
Figure 5.3: *spike* and *pit* superpixel patterns graphs at scale 1 for a selected mass ROI.

(a) Mass ROI   (b) p_0 (*spike*)   (c) p_{255} (*pit*
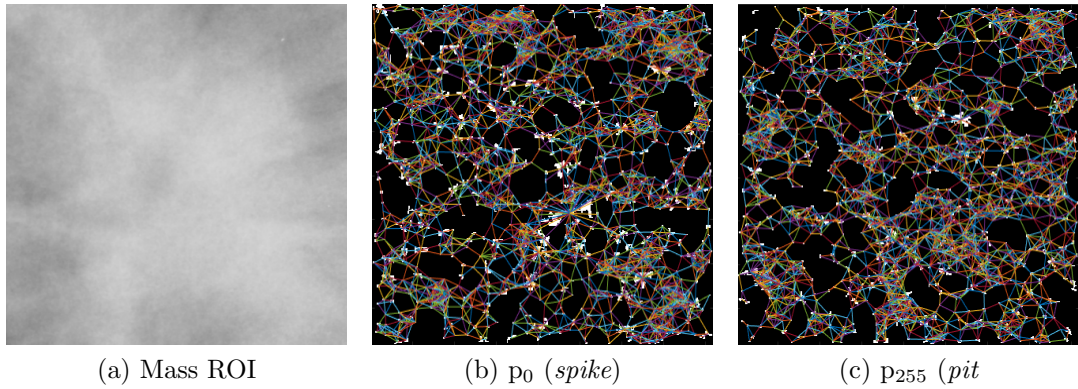
Figure 5.4: *spike* and *pit* superpixel patterns graphs at scale 2 for a selected mass ROI.

5.3.3 Topological Feature Extraction

The topological properties of graphs can be captured by various graph metrics (Chen et al. 2015). In this study, graph theoretical features such as average vertex degree, cluster coefficient, percentage of isolated nodes and giant connected component ratio are investigated. These properties are defined below.

Let $G(V, E)$ be a graph where V is the vertex set and E is the edge set. The cardinality of V and E is represented by $|V|$ and $|E|$, respectively.

1. **Average Vertex Degree:** The *degree of a vertex k* in a graph $G(V, E)$, represented by $d(k)$, is the number of edges incident on that vertex. The average vertex degree is calculated as $\sum_{k \in V} d(k)/|V|$. This feature describes the average relatedness between a structured superpixel and its neighboring structured superpixels in the ROI.

2. **Average Clustering Coefficient:** The *clustering coefficient* of a vertex k in a graph $G(V, E)$, represented by $c(k)$, is the ratio between number of edges adjacent to k and the number of all possible edges between $k's$ neighbors. The average clustering coefficient is defined by $\sum_{k \in V} c(k)/|V|$. This feature indicates how concentrated a superpixel's neighborhood is or, more precisely, how close it is to being a complete graph (a clique).

3. **Percentage of Isolated Vertices:** An *isolated vertex* of a graph $G(V, E)$ is a vertex with degree of zero. The percentage of isolated vertices is the ratio of the number of isolated vertices to the total number of vertices in G.

4. **Giant Connected Component Ratio:** The *giant connected component* of a graph $G(V, E)$ is the largest set of vertices that are reachable from any other. The ratio of the number of vertices in the giant connected component to the total number of vertices in $|V|$ is called giant connected component ratio. This feature shows the percentage of superpixels in the giant connected component with respect to all superpixels in an ROI.

5.3.4 Feature Selection

In this chapter, the maximum dimensionality of the feature space is 72 (9 structured superpixel patterns, 2 scales, 4 graph features = $9 \times 2 \times 4 = 72$). Feature selection was conducted with DDSM dataset1 similar to Chapters 3 and 4. However, in this chapter, exhaustive search was limited to three structured superpixel patterns graph combination features, due to computational complexity ($2^{72} - 1$ feature combinations) and observations from previous experiments (in both Chapters 3 and 4, best performance was achieved with three pattern combination, further increase in pattern combinations did not increase the classification accuracy). The three structured superpixel patterns graph combination resulted in 5805 feature combinations as follows. Individual structured superpixel pattern graphs resulted in 405 feature combinations (9 graphs corresponding to nine superpixel patterns, 2 scales (3 combinations) and 4 graph features (15 combinations) = $9 \times 3 \times 15 = 405$). Two graph combination, resulted in 1620 feature combinations (9 graphs in combination of 2 (36 combinations), 3 scale combinations and 15 graph features combination = $36 \times 3 \times 15 = 1620$) and three graph combinations resulted in 3780 feature combinations (9 graphs in combination of 3 (84 combination), 3 scale combinations and 15 graph features combination =

$84 \times 3 \times 15 = 3780$). Similar with the previous two chapters, learning algorithm (LDA), performance evaluation method (AUC) and validation method (ten independent runs of 5-fold cross validation) were used to analyse the effectiveness of different graph features extracted from structured macro-patterns for the separation of cancerous and normal regions. These details are summarized in Table 5.1.

Table 5.1: Summary of feature selection experiments.

Dataset	DDSM
Sample Size	93 ROIs (41 mass ROIs and 52 normal ROIs)
Classifier	LDA
Feature Selection	Exhaustive search and AUC
Validation method	5-fold cross validation
No. of repetitions	10
Total no. of features/combinations	72 features and $2^{72} - 1$ feature combinations
No. of feature combinations considered for analysis	5805 feature combinations
• Individual structured superpixel pattern graphs	405 feature combinations
• Two graph combination	1620 feature combinations
• Three graph combinations	3780 feature combinations

While analysing structured superpixel patterns graph combinations up to 3 (total 5805 feature combinations), the highest AUC score of 0.935 was obtained for patterns $\{p_0,\ p_1,\ p_{255}\}$ for scale 1 and 2 combination, and average vertex degree, average clustering coefficient and percentage of isolated vertices graph feature combination. $\{p_0,\ p_1,\ p_{255}\}$ means graph features extracted from p_0, p_1 and p_{255} combined. In addition, out of 5805 feature combinations, it was found that 9 feature combinations generated from $\{p_0,\ p_1,\ p_{255}\}$ were not statistically different in performance from the highest performing feature set. Table 5.2 shows the details. As the highest performed and all other nine feature combinations were generated from $\{p_0,\ p_1,\ p_{255}\}$ as shown in Table 5.2, the graph features generated from $\{p_0,\ p_1,\ p_{255}\}$ were selected for further analysis.

5.4 Experimental Results

Similar to the previous two chapters (see Section 3.6), the ROI classification power of the selected features was evaluated on the DDSM database (301 ROIs - all 5 datasets) and the unseen BSSA database (224 ROIs - all 5 datasets). Similarly, ten independent runs of 5-fold cross validation were applied for each of the five data sets of BSSA and DDSM to ensure stable classification results. The reported results are averages of 10 runs over 5 balanced data sets.

Table 5.3 shows the performance comparison for DDSM and BSSA for graph features generated from $\{p_0,\ p_1,\ p_{255}\}$ superpixel patterns and using LDA. Col-

Table 5.2: Best performed DDSM classification results from 10 runs of 5-fold cross validation for dataset1 using LDA. Column 1 represents the scale 1 and 2, column 2 represents the pattern combinations, column 3 represents the graph features average vertex degree, cluster coefficient, percentage of isolated vertices and giant connected component ratio as 1, 2, 3 and 4 respectively.

Scale	Patterns	Graph Features	AUC	p-value
1 & 2	$\{p_0, p_1, p_{255}\}$	1 & 2 & 3	0.935 ± 0.021	-
1	$\{p_0, p_1, p_{255}\}$	1 & 2 & 3	0.929 ± 0.021	0.803
1 & 2	$\{p_0, p_1, p_{255}\}$	1 & 3	0.927 ± 0.011	0.684
1	$\{p_0, p_1, p_{255}\}$	1 & 2 & 3 & 4	0.925 ± 0.021	0.542
2	$\{p_0, p_1, p_{255}\}$	1 & 3	0.924 ± 0.015	0.395
2	$\{p_0, p_1, p_{255}\}$	1 & 2 & 3	0.920 ± 0.012	0.213
1 & 2	$\{p_0, p_1, p_{255}\}$	1 & 2	0.919 ± 0.017	0.194
2	$\{p_0, p_1, p_{255}\}$	1 & 2 & 3	0.915 ± 0.013	0.151
2	$\{p_0, p_1, p_{255}\}$	1 & 2 & 3 & 4	0.911 ± 0.017	0.104
2	$\{p_0, p_1, p_{255}\}$	1 & 3 & 4	0.910 ± 0.012	0.066

umn 1 represents the graph features: average vertex degree, cluster coefficient, percentage of isolated vertices and giant connected component ratio as 1, 2, 3 and 4, respectively and their combinations. For example, when considering individual graph features (1, 2, 3 and 4), having one feature for p_0 (corresponding to each graph feature), 2 features for $\{p_0, p_1\}$ (individual feature for p_0 and p_1) and 3 features for $\{p_0, p_1, p_{255}\}$ } (individual feature generated for p_0, p_1 and p_{255}).

For DDSM, when considering individual graph features (1, 2, 3 and 4) generated from individual scales (1 and 2), average vertex degree feature generated from scale 2 obtained the highest AUC score of 0.875. Combining the information from scale 1 and 2 did not result in increased classifier performance. With two graph features and individual scales, combination of average vertex degree and percentage of isolated vertices features obtained the highest AUC score of 0.918 for scale 2. Combining the information from two scales the performance has increased to 0.926. With 3 graph features, combination of average vertex degree, cluster coefficient and percentage of isolated vertices for scale 1 achieved the highest AUC score of 0.928 and combining information from two scales increased the performance to 0.934. Similarly for BSSA, with single graph feature, average vertex degree feature generated from scale 2 achieved the highest AUC score of 0.734. Combining the information from scale 1 and 2 did not increase the performance. With 2 graph features, average vertex degree and giant connected component ratio generated from scale 1 achieved the highest AUC score of 0.808. Combining the information from scale 1 and 2 and further increase in

graph features did not result in increased performance. The highest scores are highlighted in bold in the table. SVM-L classifier results for DDSM and BSSA are shown in Table 5.4. The highest scores are highlighted in bold in the table.

Table 5.3: Comparison of AUC scores obtained for different graph features and their combinations generated from $\{p_0, p_1, p_{255}\}$ superpixel patterns graph for DDSM and BSSA mass classification using LDA classifier for scale 1, scale 2 and scale 1 and 2 combination. Column 1 represents the graph features average vertex degree, cluster coefficient, percentage of isolated vertices and giant connected component ratio as 1, 2, 3 and 4 respectively and their combinations.

	DDSM - LDA								
	Scale 1			Scale 2			Scale 1, 2		
Graph features	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.837	0.661	0.838	**0.875**	**0.766**	**0.843**	0.873	0.741	0.853
2	0.742	0.540	0.816	0.705	0.416	0.838	0.732	0.547	0.804
3	0.847	0.640	0.862	0.714	0.438	0.870	0.841	0.626	0.858
4	0.745	0.559	0.764	0.810	0.538	0.891	0.816	0.624	0.830
1&2	0.884	0.775	0.829	0.911	0.844	0.843	0.914	0.855	0.835
1&3	0.892	0.777	0.825	**0.918**	**0.832**	**0.855**	**0.926**	**0.849**	**0.870**
1&4	0.812	0.671	0.820	0.896	0.797	0.857	0.885	0.752	0.847
2&3	0.843	0.689	0.851	0.734	0.502	0.833	0.826	0.672	0.843
2&4	0.771	0.646	0.776	0.780	0.564	0.866	0.797	0.666	0.801
3&4	0.831	0.674	0.829	0.815	0.594	0.873	0.841	0.667	0.846
1&2&3	**0.928**	**0.866**	**0.852**	0.918	0.878	0.843	**0.934**	**0.894**	**0.852**
1&2&4	0.868	0.731	0.842	0.903	0.868	0.838	0.899	0.828	0.820
1&3&4	0.877	0.757	0.817	0.909	0.805	0.863	0.900	0.805	0.863
2&3&4	0.841	0.718	0.833	0.792	0.608	0.865	0.837	0.732	0.826
1&2&3&4	0.918	0.836	0.838	0.909	0.848	0.845	0.905	0.847	0.848
	BSSA - LDA								
	Scale 1			Scale 2			Scale 1, 2		
Graph features	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.710	0.461	0.770	**0.734**	**0.543**	**0.769**	0.734	0.590	0.765
2	0.632	0.354	0.788	0.453	0.123	0.889	0.688	0.465	0.775
3	0.623	0.288	0.842	0.565	0.161	0.944	0.602	0.348	0.825
4	0.686	0.299	0.910	0.470	0.068	0.916	0.633	0.320	0.817
1&2	0.686	0.508	0.751	0.720	0.547	0.769	0.692	0.526	0.770
1&3	0.651	0.441	0.735	0.699	0.471	0.764	0.643	0.477	0.718
1&4	**0.808**	**0.609**	**0.827**	0.707	0.581	0.742	0.785	0.615	0.800
2&3	0.609	0.373	0.778	0.530	0.261	0.826	0.629	0.447	0.749
2&4	0.779	0.563	0.832	0.424	0.170	0.803	0.763	0.580	0.805
3&4	0.737	0.448	0.842	0.502	0.184	0.885	0.712	0.512	0.791
1&2&3	0.631	0.468	0.720	0.679	0.495	0.762	0.634	0.466	0.721
1&2&4	0.782	0.617	0.795	0.676	0.529	0.750	0.744	0.584	0.776
1&3&4	0.804	0.639	0.805	0.678	0.494	0.744	0.719	0.544	0.783
2&3&4	0.731	0.534	0.802	0.515	0.270	0.788	0.721	0.561	0.781
1&2&3&4	0.772	0.607	0.784	0.655	0.477	0.748	0.689	0.536	0.758

Table 5.4: Comparison of AUC scores obtained for different graph features and their combinations generated from $\{p_0, p_1, p_{255}\}$ superpixel patterns graph for DDSM and BSSA mass classification using SVM-L classifier for scale 1, scale 2 and scale 1 and 2 combination. Column 1 represents the graph features average vertex degree, cluster coefficient, percentage of isolated vertices and giant connected component ratio as 1, 2, 3 and 4 respectively and their combinations.

DDSM - SVM-L								
	Scale 1			Scale 2			Scale 1, 2	
Graph features	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.712	0.463	0.692	**0.769**	**0.585**	**0.731**	**0.771**	**0.707**	**0.731**
2	0.538	0.171	0.827	0.508	0.049	0.885	0.557	0.366	0.692
3	0.772	0.415	0.904	0.557	0.171	0.962	0.811	0.634	0.827
4	0.642	0.463	0.712	0.777	0.463	0.865	0.747	0.512	0.731
1&2	0.733	0.512	0.769	0.830	0.780	0.769	0.823	0.854	0.769
1&3	**0.872**	**0.756**	**0.750**	0.855	0.854	0.731	**0.898**	**0.902**	**0.788**
1&4	0.540	0.463	0.654	0.857	0.854	0.712	0.801	0.683	0.750
2&3	0.659	0.537	0.750	0.665	0.341	0.808	0.730	0.634	0.769
2&4	0.668	0.488	0.750	0.776	0.463	0.846	0.743	0.659	0.769
3&4	0.749	0.659	0.750	0.726	0.366	0.846	0.828	0.707	0.788
1&2&3	0.889	0.829	0.788	0.845	0.756	0.731	**0.927**	**0.902**	**0.846**
1&2&4	0.718	0.634	0.788	0.819	0.805	0.750	0.806	0.780	0.750
1&3&4	0.807	0.707	0.712	0.812	0.707	0.712	0.856	0.707	0.731
2&3&4	0.746	0.610	0.750	0.709	0.415	0.788	0.730	0.707	0.769
1&2&3&4	0.877	0.732	0.788	0.876	0.732	0.769	0.873	0.732	0.712
BSSA - SVM-L								
	Scale 1			Scale 2			Scale 1, 2	
Graph features	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
1	0.670	0.214	0.905	**0.772**	**0.517**	**0.878**	**0.775**	**0.393**	**0.905**
2	0.500	0.143	0.881	0.448	0.000	1.000	0.641	0.310	0.805
3	0.653	0.071	0.857	0.591	0.107	0.881	0.659	0.241	0.878
4	0.723	0.429	0.905	0.669	0.034	0.951	0.773	0.517	0.878
1&2	0.648	0.379	0.780	**0.783**	**0.571**	**0.881**	0.731	0.517	0.829
1&3	0.607	0.250	0.881	0.653	0.393	0.833	0.667	0.357	0.857
1&4	0.866	0.690	0.805	0.699	0.357	0.881	**0.835**	**0.607**	**0.857**
2&3	0.594	0.276	0.902	0.643	0.310	0.878	0.679	0.464	0.881
2&4	0.748	0.643	0.810	0.478	0.138	0.878	0.795	0.607	0.810
3&4	0.715	0.500	0.762	0.555	0.103	0.854	0.724	0.464	0.762
1&2&3	0.589	0.379	0.805	0.749	0.517	0.829	0.772	0.586	0.878
1&2&4	0.788	0.679	0.810	0.753	0.571	0.857	**0.868**	**0.828**	**0.805**
1&3&4	0.787	0.571	0.833	0.739	0.393	0.833	0.786	0.621	0.805
2&3&4	0.858	0.690	0.829	0.628	0.179	0.881	0.784	0.621	0.854
1&2&3&4	0.759	0.607	0.786	0.640	0.429	0.762	0.803	0.724	0.805

5.5 Comparison with recent state-of-the art methods

The results from the proposed approaches (Chapter 3, 4 and 5) were compared with recent state-of-the art mass classification found in literature. Table 5.5 shows the comparison results. It is important for a fair comparison to use images from the same database and the same ROI extraction method. Therefore, the proposed approaches were compared with studies that used images from DDSM, manually cropped ROIs, and classification of mass and normal regions. It is also worth emphasizing that the proposed techniques were applied to dense ROIs, which is a particularly difficult and challenging task in mass classification, whereas the data used in the other studies have only a small percentage of ROIs located in the dense areas of mammograms. Despite this, the proposed mass classification method performed comparably or better with a very small number of features. All other studies used at least a couple of hundred features to get such high classification scores.

The method we presented in Chapter 3 achieved an AUC score of 0.96 (see Table 3.4 on page 53 $\{f_0, f_1, f_{255}\}$ with radius 1 and 4 information). The method presented in Chapter 4 also achieved an AUC score of 0.96 (see Table 4.3 on page 69 $\{f_0, f_1, f_{255}\}$ with level 1 and 3 information). The method presented in this chapter achieved an AUC score of 0.92 (see Table 5.3 on page 83 $\{ p_0, p_1, p_{255} \}$ for scale 2 and average vertex degree and percentage of isolated vertices combination). This makes the proposed approaches a very promising method towards building a robust mass classification system.

Table 5.5: Performance comparison of the proposed approaches with the recent state-of-the art mass classification techniques found in literature.

Works	Database	ROI selection	Dense only	Technique	No.of features	AUC
Gargouri et al. (2012)	DDSM	Manual	No	GLLD with ANN	1000+	0.95
Hussain (2013)	DDSM	Manual	No	MSWLD with SVM	337	0.99 ± 0.003
Nascimento et al. (2013)	DDSM	Manual	No	Wavelets and Polynomial classifier	6000	0.98 ± 0.030
Junior et al. (2009)	DDSM	Manual	No	Geary's coefficient with SVM	240	0.94
Hussain et al. (2014)	DDSM	Manual	No	Gabor features with SVM	2000+	0.96 ± 0.021
Choi & Ro (2012)	DDSM	Automatic	No	LBP with SVM	255	0.92 ± 0.026
Llado et al. (2009)	DDSM	Manual	No	LBP with SVM	2000+	0.94 ± 0.020
Proposed - Chapter 3	DDSM	Manual	Yes	**structured micropatterns with LDA**	**6**	$\mathbf{0.96 \pm 0.005}$
Proposed - Chapter 4	DDSM	Manual	Yes	**SRM-SLBP with LDA**	**6**	$\mathbf{0.96 \pm 0.005}$
Proposed - Chapter 5	DDSM	Manual	Yes	**Superpixel patterns graph with LDA**	**6**	$\mathbf{0.92 \pm 0.022}$

However, the proposed approaches (Chapters 3, 4 and 5) have considered a smaller number of normal ROIs (manually cropped ones) when compared to the automatically generated ones in the real CAD. To overcome this limitation the

features selected from the proposed approaches are evaluated using a larger set of automatically generated ROIs in Chapter 7.

5.6   Discussion and Conclusions

In this chapter, we have presented a novel approach for classifying dense ROIs in mammograms based on topology analysis of structured superpixel patterns generated using the SLBP technique. The SRM technique was used to generate the superpixels, as SRM superpixels are statistically homogeneous in terms of pixel gray level values and show high classification potential, based on the results from the previous study (Chapter 4). Superpixel size $\approx(3 \times 3)$ was selected in this study as it has shown better performance than superpixel sizes $\approx(4 \times 4)$ and $\approx(5 \times 5)$ (details in Chapter 4). Multiscale morphology was used to analyze the spatial relationships between the individual superpixel patterns. In this study we have focused on four clinically recognizable features that were extracted from the superpixel patterns graphs for mass classification. Even though all features related to the relative closeness of individual superpixel patterns in the ROI, each feature gives different information in describing the topological structure of superpixel pattern clusters. Alike average vertex degree gives information of interior connectivity within the superpixel pattern, which is a local feature, whereas giant connected component ratio gives the percentage of the largest connected component size over the size of the entire graph resulting from dilating each superpixel, is a global measure.

Similar with the results from the two previous chapters, *spike* (p_0 or p_1) and *pit* (p_{255}) combination gave the best results. No biological explanation was found in the literature for this observation.

Overall results show that the proposed approach can identify meaningful superpixel patterns present within mass-like regions. The topology of superpixel patterns, captured using spatial connectivity graphs, can reveal significant differences between cancerous and healthy areas of breasts.

Finally, it is worth mentioning that the proposed technique does not involve any parameter tuning and by its nature as superpixel based method is robust to noise present in images and is computationally efficient since the graph was built on superpixels instead of pixels.

In the future, we plan to investigate other properties of connectivity graphs generated from subsets of the nine superpixel local binary patterns. Similarity

measures between the superpixel patterns graphs could be investigated. In addition to LDA and SVM, classifiers such as random forests, ANN, and ensemble classifiers could also be explored. In this study we have built the graph on structured superpixel patterns generated using SLBP technique. Additional texture representation techniques could be investigated for graph study.

Chapter 6

Improving Breast Mass Segmentation in Local Dense Background

In an automatic mass localisation system, the ROIs are selected by mass segmentation algorithm followed by a method for prescreening of suspicious ROIs. The automatically extracted suspicious ROIs may have overlapping mass and normal tissues which may reduce the accuracy of the algorithm for finding efficient texture descriptors for mass ROIs. Hence for dense ROI texture analysis and classification in Chapters 3, 4 and 5, manually cropped ROIs were used to find the efficient texture descriptors. However the selected features are evaluated by the automatically extracted ROIs in the mass localisation stage (described in Chapter 7).

Accurate mass candidate segmentation is an important step for the subsequent mass localisation process and, from the literature, it is clear that segmentation of masses in dense background is a challenging task (Bajger et al. 2009, Ma et al. 2009, Ho & Lam 2003, Obenauer et al. 2006). Hence in this chapter, two methods for improving the breast mass segmentation in local dense background are presented. The first method analyses the effects of image enhancement on mass segmentation. In this method, a self-adjusted mammogram contrast enhancement solution called adaptive CLAHE is developed, aiming to improve mass segmentation. In the second method, an optimization algorithm for automatically tuning segmentation of mammograms by SRM technique is proposed, to improve the mass segmentation. Section 6.1 describes the image dataset used for conducting experiments in this chapter. Section 6.2 describes the adaptive CLAHE which au-

tomatically selects the clip limit and block size for individual images. Section 6.3 describes the self-adjusted SRM which automatically selects the best Q value for optimal segmentation. Section 6.4 provides the discussion and summary.

6.1 Dataset

As the focus of this thesis is to find efficient descriptors for masses in dense regions, the previous three chapters used only mammograms with masses localized in dense background. However, in this chapter, the dataset is extended with images of masses in nondense background to investigate performance of the proposed methods in masses in nondense background.

For experiments in this chapter, all the malignant mammograms with masses localized in dense background from the DDSM database (41) and the BSSA database (29), and 41 mammograms from DDSM having masses in nondense background were selected.

Experiments conducted for this chapter used downsampled images (by a factor of 8) to reduce the processing time. The effect of downsampling is discussed in last paragraph of Section 6.2.6.

6.2 Optimized Image Enhancement for Improving Mass Segmentation

Studies have shown that image enhancement can increase the contrast between malignant and normal tissue in dense breast and improve mass localisation/ detection (Jo et al. 2013, Pandey et al. 2012, Singh & Bovis 2005). Jo et al. (2013) applied Adaptive Histogram Equalization (AHE) to increase the detection rate of cancer in dense breast. Pandey et al. (2012) used Adaptive Volterra Filter to improve the contrast of mammographic masses from the surrounding tissues. Singh & Bovis (2005) proposed a set of metrics (Distribution Separation Measure, Target-to-Background Contrast Enhancement Measurement Based on Entropy and Combined Enhancement Measure) to measure the quality of the image enhancement of mammographic images in a CAD for finding masses using machine learning techniques. Based on these metrics, performance of different contrast enhancement techniques was evaluated. According to their study, a good enhancement method may greatly improve the segmentation accuracy of

mass detection in dense breasts compared to the segmentation obtained from an unenhanced original image. Their experimental results showed that image enhancement has a significant influence on image segmentation and optimizing enhancement on a per image basis, giving better results compared to using the same method for all images. Choosing a single best technique for image enhancement is a difficult task and the performance of the enhancement is often evaluated based on the performance of the subsequent segmentation performance (Gonzalez & Woods 1992). Performance of the proposed adaptive CLAHE is also evaluated using the subsequent segmentation performance.

6.2.1 CLAHE

Pisano et al. (2000) tested different image enhancement techniques: Manual Intensity Windowing, Histogram-based Intensity Windowing, Mixture-Model Intensity Windowing, Contrast-Limited Adaptive Histogram Equalization (CLAHE), unsharp masking, peripheral equalization, and Trex processing for digital mammography to check how these algorithms may affect the ability of radiologists to interpret the images. According to their study, the CLAHE method improves the detection of simulated spiculations in dense mammograms compared to others. The CLAHE algorithm is extensively used by various researchers for CAD applications in medical imaging because of its efficiency and straightforward implementation (Sundaram et al. 2011, Rahmati et al. 2010, Shelda & Ravishankar 2013, Wu et al. 2013, Maitra et al. 2012).

Sundaram et al. (2011) used CLAHE based contrast enhancement for mammograms. Rahmati et al. (2010) used fuzzy CLAHE as a preprocessing filter to eliminate the noise and intensity inhomogeneities in mammograms to improve segmentation of masses. Wu et al. (2013) adopted CLAHE to enhance the high frequency subbands coefficients in-order to enhance the features and image contrast. Maitra et al. (2012) also used CLAHE as a preprocessing technique for digital mammograms. CLAHE can effectively remove the noise and enhance the local features, edges and image contrast without losing any relevant information in the original mammogram image (Wu et al. 2013, Maitra et al. 2012). However, the performance of the standard CLAHE technique depends on two key parameters, clip limit (c) and block size (b), and these parameter values were heuristically chosen by users (Sundaram et al. 2011, Rahmati et al. 2010, Shelda & Ravishankar 2013, Wu et al. 2013, Maitra et al. 2012).

In this chapter, an adaptive CLAHE based on entropy is proposed to improve

the segmentation of masses in local dense background. As mentioned previously, standard CLAHE heavily depends on the clip-limit c and block size b, which are user defined. In our approach, these two parameters are determined adaptively for each image using an optimization algorithm that uses a measure of entropy. To analyse the effect of the proposed enhancement method on mass segmentation, FCM clustering is used to produce the mass candidates. Figure 6.1 shows the block diagram for mass segmentation.
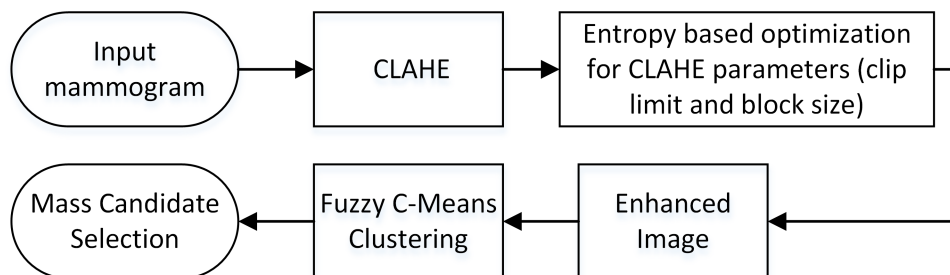
Figure 6.1: Flow chart for mass segmentation

6.2.2 Adaptive CLAHE based on Entropy

6.2.2.1 An Overview

The standard CLAHE algorithm is a modification of AHE (Pizer et al. 1987), which limits the amplification by clipping the histogram at a user-defined value called the clip limit. The clip limit determines how much noise in the histogram should be smoothed and, hence, how much contrast should be enhanced. CLAHE equally redistributes the histogram above the clip limit among all the histogram bins. The histogram can have different distributions such as uniform, exponential, Rayleigh etc. A uniform probability density distribution does not help in the mass localisation in dense regions, as it simultaneously distributes the dynamic range between background and foreground. Therefore, Rayleigh distribution which is a nonuniform distribution function, is used in this study. Finally, a cumulative distribution function is determined for the gray scale mapping where the mapping at each pixel is interpolated using bi-linear interpolation of the neighboring pixels.

Even though CLAHE with Rayleigh distribution gives good contrast enhancement for mammogram images, it heavily depends on the clip limit and block size. These parameter values are set up by the user. When a user determines inappropriate parameter values, the results of the CLAHE may be worse than that of the original image. Figures 6.2 shows an example of CLAHE with different clip limits

91

and block sizes. In un-enhanced image mass located in a local dense background is hardly visible and the boundaries are extremely hard to detect (See Figure 6.5 (a)). This can cause failure of mass localisation in the original image due to under segmentation, while CLAHE with proper parameter settings can provide better noise removal and enhanced contrast between mass and background, which can help in proper mass segmentation (See Figure 6.5 bottom row).

(a) (b) (c)

(d) (e) (f)
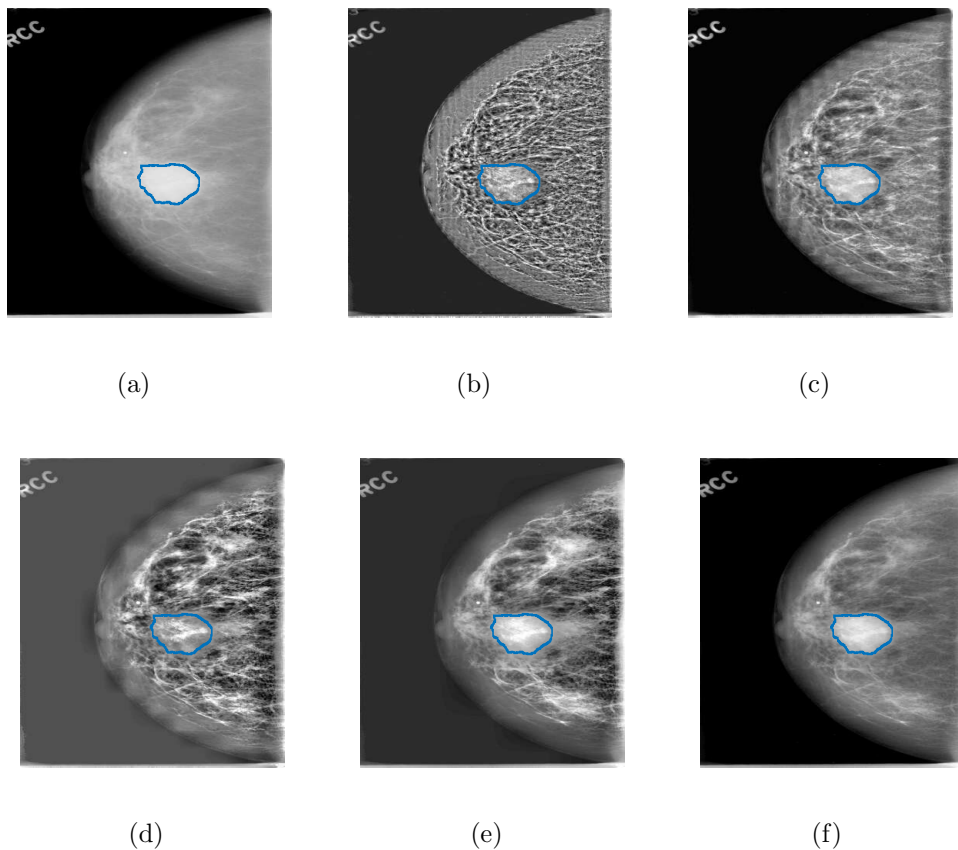
Figure 6.2: Example of CLAHE application with different clip limit and block size on a mammogram from DDSM (a) Original image (b) CLAHE (c=0.1, b=8 × 8) (c) CLAHE (c=0.02, b=16 × 16) (d) CLAHE (c=0.3, b=32 × 32) (e) CLAHE (c=0.1, b=64 × 64) (f) CLAHE (c=0.008, b=64 × 64)

The parameters clip limit and block size controls the contrast of the image and hence the quality of enhancement. In most applications of the CLAHE algorithm, clip limit and block size are fixed empirically for a class of images with results that are far from the optimal for some members of the class. In this chapter, clip limit and block size are adjusted automatically for each image by optimizing a measure of entropy as described in the next section.

6.2.2.2 Entropy of Enhanced Image

Adjusting parameters values without human intervention is a difficult task in image processing. This is because automatic image enhancement requires specifying an objective criterion for enhancement and this objective criterion should adjust the quality of the image for the particular task. This study of automatically tuning the clip limit and block size of CLAHE uses an unsupervised evaluation criterion based on information theory (entropy) that adjusts the quality of the enhanced image without any prior knowledge. The entropy value predicts/reveals the information contained in an image and, hence, widely used in optimizing image enhancement techniques (Niu et al. 2016, Panetta et al. 2008, Wan et al. 2018, Shelda & Ravishankar 2013). Based on the image histogram, the entropy value is calculated on the enhanced image I_e as given below (Shannon 1948, Pun 1980)

$$H(I_e) = -\sum_{i=0}^{L-1} h_i log_2 h_i, \tag{6.1}$$

where h_i is the probability occurrence of the intensity value in the enhanced gray image I_e and intensity values/gray levels $i \in \{0, 1, \cdots L-1\}$.

6.2.2.3 Optimization Algorithm

To determine optimal values for the clip limit (c) and block size (b) the following algorithm was used. Every mammogram was enhanced using CLAHE with a limited range of values of c and b. For each value of c and b, the entropy of the enhanced image was computed using Equation 6.1. The value of c and b resulting in the maximum difference in entropy were selected as optimal parameter values for the enhanced image. The reasonable range (estimated experimentally) for c was selected as $0.001 : 0.001 : 0.02$, and block size as 16×16, 32×32, 48×48 and 64×64.

6.2.3 FCM

To demonstrate the effectiveness of automatically optimized clip-limit and block size of CLAHE, the FCM clustering algorithm was adopted. FCM clustering is
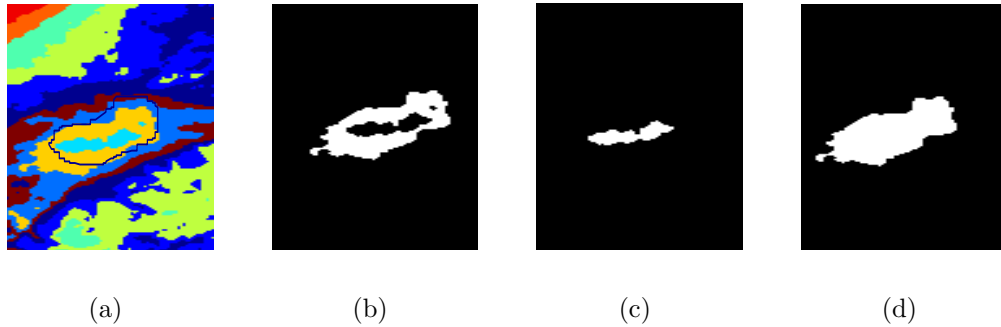
Figure 6.3: Mass candidate segmentation with morphological filling (a) ROI showing the location of the mass (b) Outer component of the selected mass candidate (c) Inner component (d) Single mass component after applying morphological filling

one of the most popular algorithms for segmentation. FCM uses iterative optimization of an objective function based on weighted similarity measures between the pixels in the image and each cluster center (Rangayyan 2005). The FCM algorithm was described in detail in Section 2.2.3. In this chapter, for the FCM algorithm, 10 clusters were used to determine the mass candidate. Ten clusters were chosen based on our previous experience on the mini-MIAS database.

6.2.4 Mass Candidate Selection

Masses are usually hyper-dense with respect to the background with core parts having high intensity values that tend to decrease as the distance to core parts increases (Rangayyan 2005). This property is preserved with the proposed approach. This allows morphological filling to be used to get a good mass like component, as shown in Figure 6.3. The figure shows the mass in an FCM clustered image. It can be observed that the mass has an inner core with a high-intensity that decreases as the distance to core increases. This high intensity region of the mass is identified as a single core component with the outer area being another component. Morphological filling is used to obtain mass like components. The component whose centroid resides inside of the annotated region (ground truth) with the highest Dice index, described next, is accepted as the mass region.

6.2.5 Performance Measures

The performance measures used are Dice index and Hausdorff distance. They are popular similarity measures for sets. The Dice index for two sets is calculated as

follows (Dice 1945)

$$DICE(X, Y) = \frac{2 \mid X \cap Y \mid}{|X| + |Y|},$$

where X and Y are the two sets to be measured. It is calculated by simply taking twice the number of elements common to both sets divided by the total number of elements in the two sets. The Dice index value ranges between 0 and 1. A value of 0 indicates that two sets have no common elements and value 1 indicates that the segmentation result and ground truth overlap entirely.

Hausdorff distance is the maximum distance of a set to the closest point in the other set (Rote 1991). In this chapter, Hausdorff distance is used as measure of the distance between the segmented mass candidate and the ground truth. Lower value for Hausdorff distance is preferable for segmentation evaluation. Let $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$ are two sets with m and n number of elements, then the Hausdorff distance from X to Y is defined as follows:

$$h(X, Y) = \max_{x \in X} \min_{y \in Y} \|x - y\|$$

As most of the times $h(X, Y) \neq h(Y, X)$, a general definition of Hausdorff distance between two sets X to Y is defined as:

$$HD(X, Y) = \max(h(X, Y), h(Y, X))$$

6.2.6 Results

The performance of the proposed adaptive CLAHE is compared with five enhancement techniques: Adjustable HE (Arici et al. 2009), traditional Unsharp Masking (UM), neutrosophy based enhancement (Guo & Cheng 2009), standard CLAHE, adaptive clip limit (ACL)-CLAHE based on standard deviation (Abbas et al. 2013) and original image (no enhancement). For the enhancement techniques, the best parameters values were determined empirically. For mass in local dense background, the parameters values for Adjustable HE (sigma), UM (scaling factor), CLAHE (clip-limit and block-size), ACL-CLAHE based on standard deviation (block-size) and neutrosophy (alpha and beta) were 0.6, 0.7, 0.013, 64×64, 64×64, 0.85 and 0.85 and for mass in local nondense background were 0.2, 0.5, 0.01, 64×64, 64×64, 0.85 and 0.85 respectively.

Table 6.1 shows the performance comparison for mass segmentation using Dice index for both mass in local dense and local nondense background. The results demonstrate that the mass segmentation performance with the proposed adaptive CLAHE enhancement is significantly better than the other five methods. With the proposed method, 95% of the images were segmented with a high Dice index for mass in local dense background (BSSA and DDSM - dense) and 98% for mass in local nondense background (DDSM - nondense). It shows an increase of 44% for mass segmentation (Dice index greater than 0.5) in local dense background and 15% for mass segmentation in local nondense background in comparison with the original image with no enhancement for the DDSM dataset. Similar results were achieved for BSSA dataset. For both mass segmentation in local dense and nondense background, the proposed method is better compared with all others. Figure 6.4 shows plots of the proportion of the correctly segmented masses as functions of Dice index for mass in dense and nondense background. These plots also indicate that the proposed adaptive CLAHE method, using entropy performance is significantly better than original image and other commonly used enhancement methods.

Table 6.1: Performance comparison for FCM mass segmentation using Dice index for BSSA (29 images) and the DDSM set for both mass in dense background (41 images) and nondense background (41 images) for each of the seven methods. Columns 2, 3 and 4 show the percentage of the number of images whose Dice index is greater than 0.5 for BSSA - mass in dense background, DDSM - mass in dense background and DDSM - mass in nondense background respectively. The corresponding number of images is shown in brackets.

Approach	BSSA - Mass in dense background	DDSM - Mass in dense background	DDSM - Mass in nondense background
Original	55% (16/29)	51% (21/41)	83% (34/41)
Adjustable HE (Arici et al. 2009)	34% (10/29)	73% (30/41)	80% (33/41)
Unsharp Masking	55% (16/29)	61% (25/41)	88% (36/41)
Neutrosophic (Guo & Cheng 2009)	66% (19/29)	71% (29/41)	85% (35/41)
CLAHE	86% (25/29)	88% (36/41)	**98% (40/41)**
ACL-CLAHE (std) (Abbas et al. 2013)	79% (23/29)	78% (32/41)	90% (37/41)
Proposed Adaptive CLAHE	**95% (27/29)**	**95% (39/41)**	**98% (40/41)**

Table 6.2 shows the performance comparison for mass segmentation using Hausdorff Distance for both mass in local dense and local nondense background. This result is also consistent with Dice index measure and it can be concluded that adaptive CLAHE enhancement is significantly better than the other six methods.

(a) BSSA - Mass in dense background

(b) DDSM - Mass in dense background

(c) DDSM - Mass in nondense background
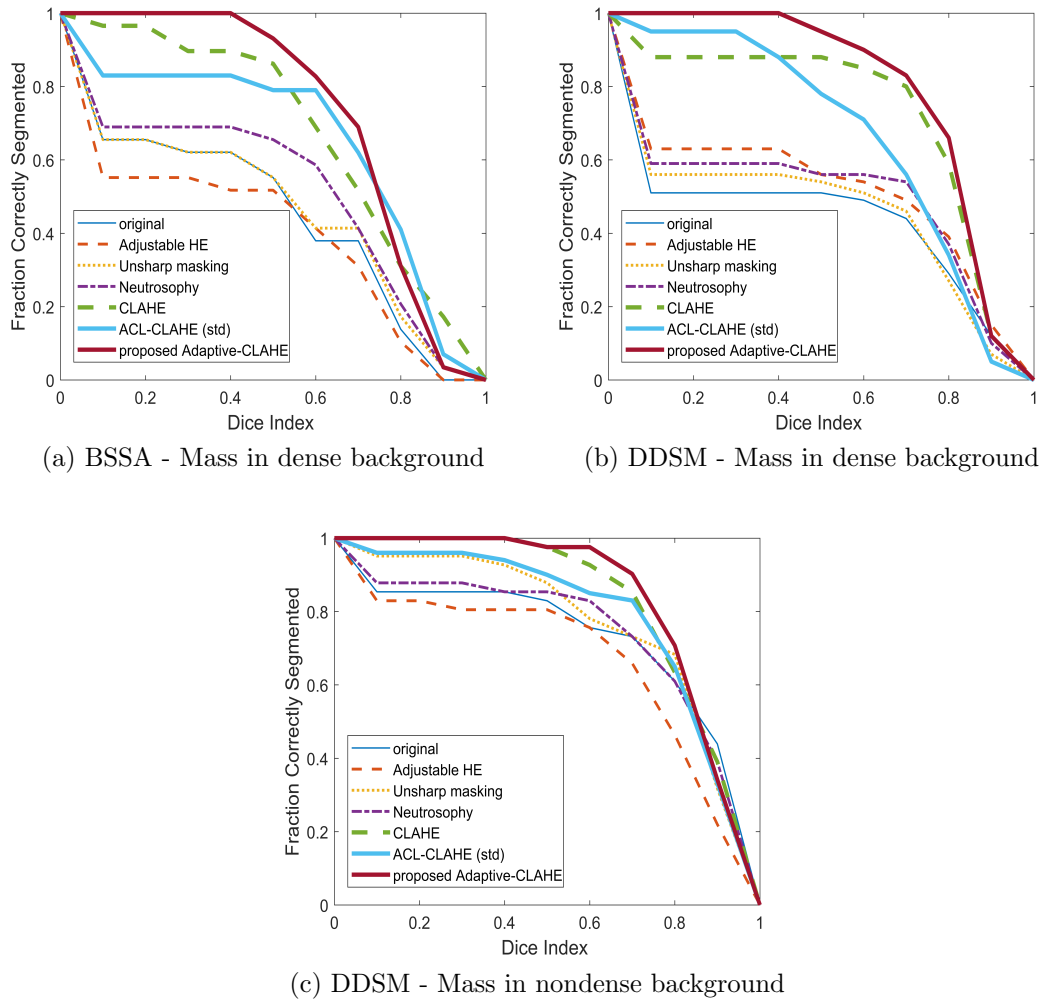
Figure 6.4: Performance comparison for FCM mass segmentation in local dense and nondense background with six different image enhancement methods and original image without any enhancement.

Figure 6.5 shows the impact of the proposed adaptive CLAHE for mass delineation in comparison with CLAHE (fixed clip limit and block size) enhanced image and the original image with no enhancement for mass in local dense background. The top row shows the process of obtaining the mass candidate for the original image, middle row shows the same for CLAHE enhanced image and the bottom row shows the proposed adaptive CLAHE enhanced image. It is evident from the original image (see Figure 6.5 top row (a)) that the mass is hardly visible and the boundaries are extremely hard to detect. Figure demonstrates that the proposed method is found to be effective in segmenting such hard cases while the mass candidate is lost in the original image as well as CLAHE enhanced image due to under segmentation. The under segmentation may be due to the high

Table 6.2: Performance comparison for FCM mass segmentation using Hausdorff Distance for BSSA (29 images) and the DDSM set for both mass in dense background (41 images) and nondense background (41 images) for each of the seven methods. Columns 2, 3 and 4 show the mean Hausdorff Distance for BSSA - mass in dense background, DDSM - mass in dense background and DDSM - mass in nondense background respectively. Standard deviation is shown in brackets.

Approach	BSSA - Mass in dense background	DDSM - Mass in dense background	DDSM - Mass in nondense background
Original (no enhancement)	5.45 (2.75)	4.00 (1.06)	4.08 (2.63)
Adjustable HE (Arici et al. 2009)	5.15 (1.73)	3.75 (0.98)	3.93 (1.83)
Unsharp Masking	5.15 (1.73)	4.04 (0.98)	3.63 (2.45)
Neutrosophic (Guo & Cheng 2009)	5.18 (2.29)	4.10 (1.07)	**3.11 (1.10)**
CLAHE	3.88 (1.13)	3.62 (0.87)	3.19 (1.30)
ACL-CLAHE (std) (Abbas et al. 2013)	3.92 (1.12)	3.92 (1.12)	3.41 (1.46)
Proposed Adaptive CLAHE	**3.90 (0.87)**	**3.15 (1.00)**	**3.11 (1.63)**

similarity in intensity values. Adaptive CLAHE is able to segment regions adequately because of the effective contrast enhancement mechanism. Morphological filling helps to overcome the over segmentation.

To understand the effect of downsampling, an experiment was conducted with original resolution images (without any enhancement) from DDSM dense dataset. It was found that after FCM segmentation, 56% of images achieved Dice index greater than 0.5 which is comparable to 55% for downsampled images (Table 6.1). In addition, the average running time was 1612 s for an original resolution image compared to 30 s for downsampled images.

6.3 Optimization of SRM Segmentation Algorithm

In addition to the method proposed in Section 6.2.2, another approach is proposed in this chapter for improving breast mass segmentation in dense regions is by optimizing the SRM segmentation algorithm without using any enhancement. This approach was motivated by the fact that any preprocessing step may suppress some important local texture information that can be used in a subsequent feature analysis step for mass localisation.
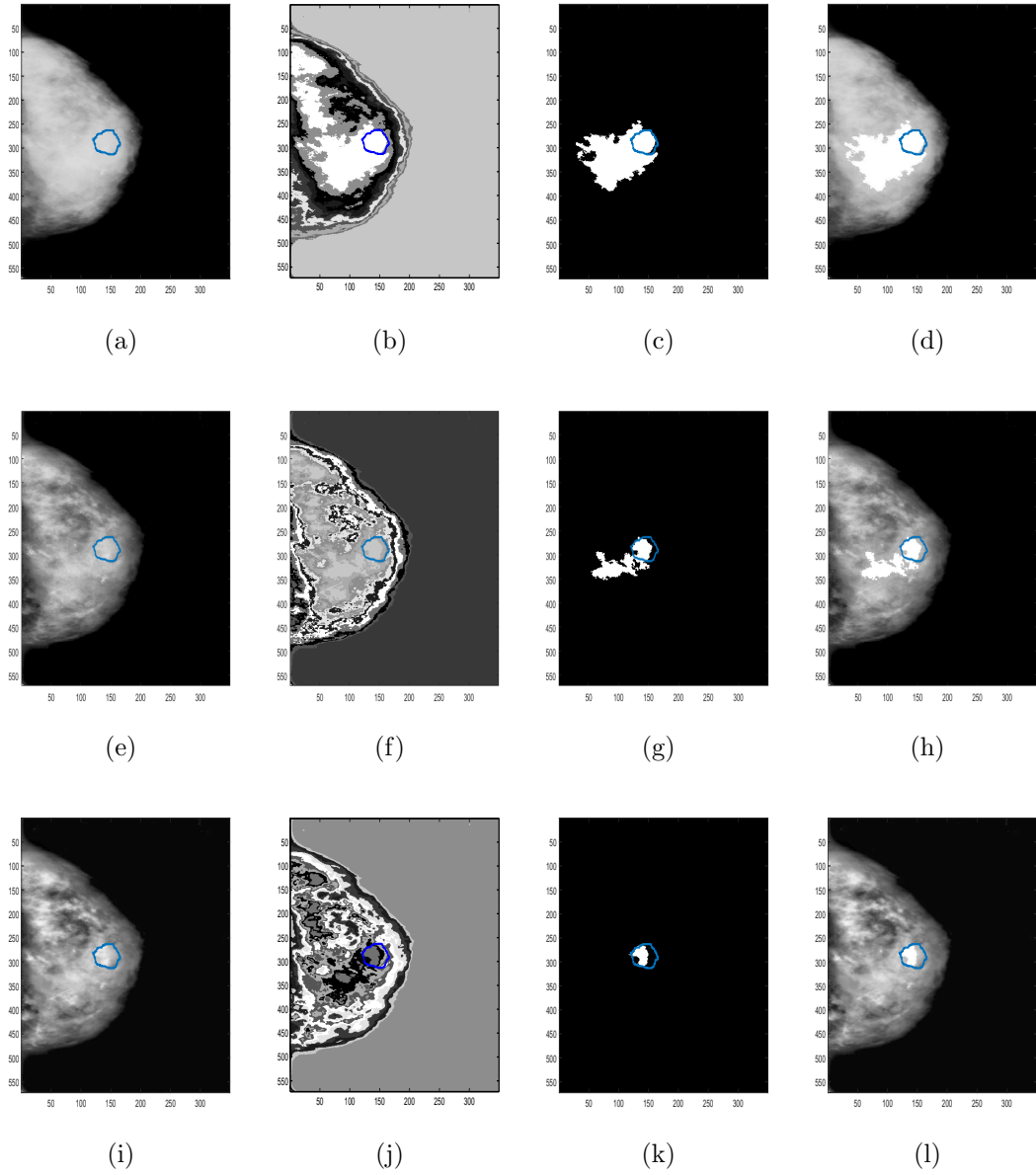
Figure 6.5: The impact of the proposed adaptive CLAHE on mass delineation, first row shows the process of obtaining the mass candidate for original image (with out any enhancement) and the second row shows the same for the CLAHE (fixed clip limit and block size) enhanced image and third row shows the proposed adaptive CLAHE enhanced image, (a), (e) and (i) original, CLAHE enhanced and proposed adaptive CLAHE enhanced image with core mass contour, respectively (b), (f) and (j) segmented image after applying FCM, (c), (g) and (k) mass candidate detected inside the ground-truth for original, CLAHE enhanced and proposed adaptive CLAHE enhanced image, respectively and finally (d), (h) and (l) mass candidate superimposed on original, CLAHE enhanced and proposed adaptive CLAHE enhanced image respectively.

6.3.1 SRM

In SRM segmentation, Q is a parameter whose value has to be set by the user. This parameter quantifies the statistical complexity of the image and hence controls the granularity of the SRM segmentation. Smaller Q values result in under segmentation and high Q values produces over segmentation. In applications using SRM technique, a fixed value of Q is selected empirically for a class of images with the result that may not be optimal for all members of the class. Hence, it is critical to select a Q value in such a way that objects of interest are well segmented. In this chapter, the Q value is selected automatically for each image by optimizing a measure of local and global image entropy.

6.3.2 Entropy of Segmented Image

It is widely accepted that some breasts appear in mammograms with very few features as they have little dense tissue that could be interpreted as mass-like, while some others contain intensity variation and many mass-like regions (Bajger et al. 2009). In both cases, a better segmentation results if a good balance is reached between the overall variation of image intensity and variation associated with mass-like regions. This can be achieved by measuring the image entropy (Susukida et al. 2008, Min et al. 2013). In (Susukida et al. 2008), a region based measure of entropy was used to automatically tune a graph based image segmentation algorithm to delineate large objects in mammograms. Our study of automatically tuning the Q value of the SRM algorithm was motivated by above mentioned work.

The approach used to optimize the SRM algorithm uses an entropy measure of the effectiveness of an image segmentation introduced in (Zhang et al. 2004). For image I, the total image entropy H is defined as

$$H(I) = H_l(I) + H_r(I). \tag{6.2}$$

In Equation (6.2), $H_l(I)$ measures the global image disorder called *layout entropy*. Usually this number increases with the number of components. It is defined by the following formula

$$H_l(I) = -\sum_{j=1}^{N} \frac{|A_j|}{|A|} \log \frac{|A_j|}{|A|},$$

where $|A|$ is the area of the whole image and $|A_j|$ is the area of the j-th segmented component and N is the number of components.

The second term $H_r(I)$, called *region entropy*, measures the uniformity within components. The region entropy decreases when the number of regions increases and is given by the formula

$$H_r(I) = \sum_{j=1}^{N} \frac{|A_j|}{|A|} H_\mu(A_j),$$

where $H_\mu(A_j)$ is the entropy of attribute μ for component A_j. In Zhang et al. (2004), luminance was used as the attribute μ. In this work, μ is the intensity value of the image pixel. Denoted by M_j the set of values associated with feature μ in component A_j and by $L_j(m)$ the number of pixels in component A_j with value m for feature μ, the entropy of component A_j is expressed as

$$H_\mu(A_j) = - \sum_{m \in M_j} \frac{L_j(m)}{|A_j|} \log \frac{L_j(m)}{|A_j|}.$$

6.3.3 Optimizing Algorithm

To optimize the value of the parameter Q, each mammogram was segmented using the SRM technique for a specified range of values of Q. Then, for each value of Q, the entropy of the segmented image was computed using the formula in (6.2). The value of Q resulting in the maximum difference in entropy was selected as optimal. The range for Q was selected empirically (50:10:300). The detailed flow-chart of the proposed technique is shown in Figure 6.6.

6.3.4 SRM Mass Candidate Selection

In the adaptive CLAHE method described in Section 6.2.2, morphological filling was used to obtain mass like components. FCM clustering produces concentric layers for segmentation with an inner high intensity mass core as one region and an outer less bright region as another region (see Figure 6.3). Hence morphological filling enabled the production of good mass candidates. Figure 6.7 shows examples of SRM mass candidate segmentation. It is clear from the figure that morphological filling does not help to produce good mass candidates. Therefore, a different approach was adopted to select the mass candidate as described below. All reasonable components that overlap the annotated region were merged
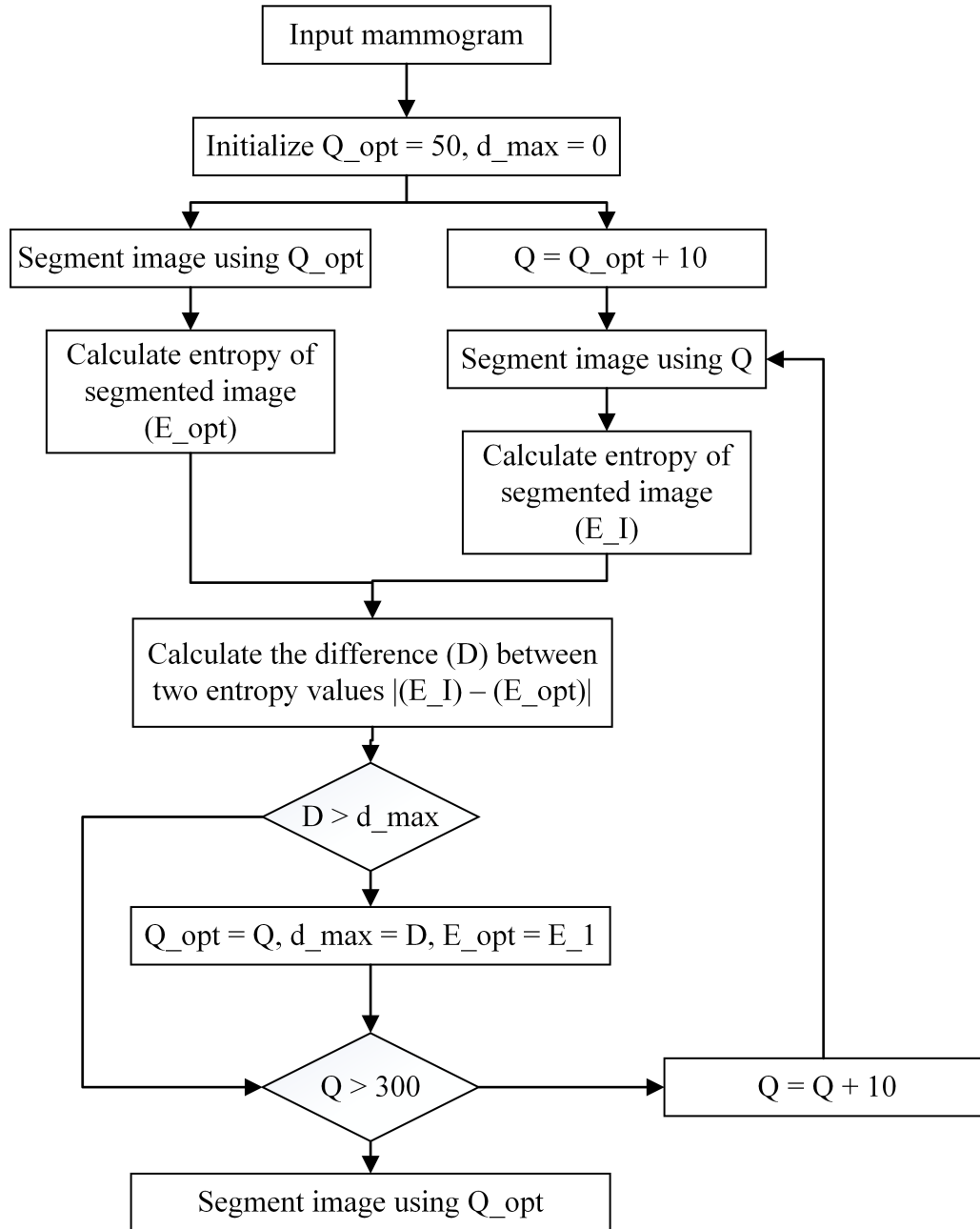
Figure 6.6: Flow-chart for proposed entropy based optimization of SRM segmentation of mammograms.

to produce mass candidates. A component with at least 60% of its area residing within the annotated region and with Dice index more than 0.2 is considered to be a reasonable component. In other words, we have used ground truth (annotated region) to merge the segmented components to analyse the potential of the approach. There should be an automatic merging criterion when applying this approach to unknown case as there will be no annotated region. This is further explained in detail in last paragraph of Section 6.4.
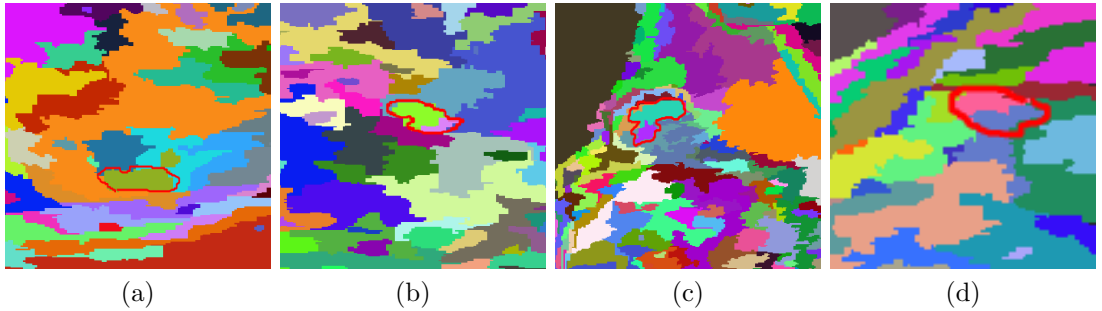
Figure 6.7: Examples of SRM mass candidate segmentation outcomes with mass contour drawn in red.

6.3.5 Results

The performance of automatically optimized SRM segmentation was compared with the FCM clustering method. For a fair comparison, the FCM method was optimized by the proposed optimization algorithm. Like the Q value in SRM, in FCM the range of cluster numbers was selected empirically (8:1:20).

Table 6.3 shows the performance comparison for mass segmentation using Dice index when merging the reasonable components as described in Section 6.3.4. The result shows that the proposed SRM parameter tuning algorithm has the potential to produce effective mass segmentation without any preprocessing (enhancement) when merging the components. The SRM outcome outperformed the FCM outcome by 44% and 17% for masses in dense background BSSA and DDSM respectively, and by 8% for masses in nondense background DDSM. Figure 6.8 show plots of the proportion of the correctly segmented masses as functions of Dice index for the union of reasonable components approach. These plots also indicate that for the original image using optimized SRM performance was significantly better compared with FCM. From this analysis, SRM was selected over FCM as the mass segmentation technique for the further investigations.

The SRM algorithm performance was further evaluated with different preprocessing (enhancement) techniques. As described in Section 6.2.5, Dice index and Hausdorff distance were used as performance measures. Table 6.4 shows the performance comparison of optimized SRM (with merging component approach) using Dice index when preprocessing was applied. For the enhancement techniques, the best parameters values were determined empirically. For mass in local dense background, the parameters values for Adjustable HE (sigma), UM (scaling factor), neutrosophy (alpha and beta), CLAHE (clip-limit and block-size), and ACL-CLAHE based on standard deviation (block-size) were 0.6, 0.2, 0.85

Table 6.3: Performance comparison of FCM and SRM for mass segmentation using Dice index for the DDSM set for both mass in dense background (41 images) and nondense background (41 images) and for BSSA (29 images) for original (unenhanced) images. Columns 2 and 3 are the percentage of the number of images whose Dice index is greater than 0.5 for FCM and SRM. The corresponding number of images is shown in brackets.

Dataset	FCM	SRM
BSSA - Mass in dense background	66% (19/29)	**100% (29/29)**
DDSM - Mass in dense background	73% (30/41)	**90% (37/41)**
DDSM - Mass in nondense background	90% (43/48)	**98% (47/48)**
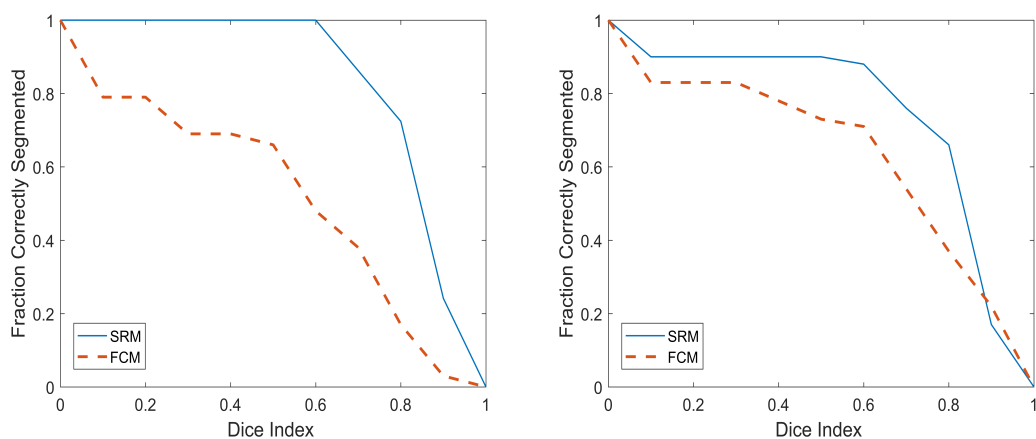
(a) BSSA - Mass in dense background   (b) DDSM - Mass in dense background
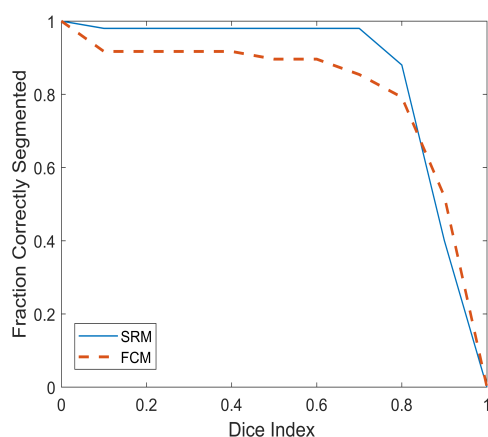
(c) DDSM - Mass in nondense background

Figure 6.8: Performance comparison for mass segmentation in local dense and nondense background using merging technique

and 0.85, 0.01 and 32×32, 32×32, and for mass in local nondense background 0.3, 0.7, 0.85 and 0.85, 0.01, 32×32, 32×32, respectively. Results show that optimized SRM without preprocessing performance is comparable with results obtained with most of the enhancement techniques. Figure 6.9 shows plots of the proportion of the correctly segmented masses as functions of Dice index for mass in dense and nondense background. These plots also indicate that the proposed optimized SRM without preprocessing, is comparable with all other commonly used enhancement methods.

Table 6.4: Performance comparison for SRM mass segmentation using Dice index for the DDSM set for both mass in dense background (41 images) and nondense background (41 images), BSSA (29 images) for each of the seven methods. Columns 2, 3 and 4 show the percentage of the number of images whose Dice index is greater than 0.5 for BSSA - mass in dense background, DDSM - mass in dense background and DDSM - mass in nondense background respectively. The corresponding number of images is shown in brackets.

Approach	BSSA - Mass in dense background	DDSM - Mass in dense background	DDSM - Mass in nondense background
Original (no enhancement)	**100% (29/29)**	90% (37/41)	**98% (40/41)**
Adjustable HE (Arici et al. 2009)	83% (24/29)	**95% (39/41)**	90% (37/41)
Unsharp Masking	69% (20/29)	90% (37/41)	95% (39/41)
Neutrosophic (Guo & Cheng 2009)	93% (27/29)	75% (30/41)	**98% (40/41)**
CLAHE	97% (28/29)	88% (36/41)	93% (38/41)
ACL-CLAHE (std) (Abbas et al. 2013)	83% (24/29)	83% (34/41)	81% (33/41)
Adaptive CLAHE (entropy)	97% (28/29)	**95% (39/41)**	**98% (40/41)**

Table 6.5 shows the SRM performance comparison using Hausdorff distance for both mass in local dense background and nondense background. This results also show that optimized SRM without preprocessing is comparable/slightly better than results obtained from enhancement techniques

6.4 Discussion and Summary

In this chapter, two methods for improving breast mass segmentation in dense background were proposed. The first was based on adaptive CLAHE followed by FCM. The performance of the proposed approach was compared with five other enhancement techniques and the original image without any preprocessing. Morphological filling was used for selecting the mass candidate, Dice index and Hausdorff distance were used as the performance measures. The experimental

(a) BSSA - Mass in dense background

(b) DDSM - Mass in dense background

(c) DDSM - Mass in nondense background
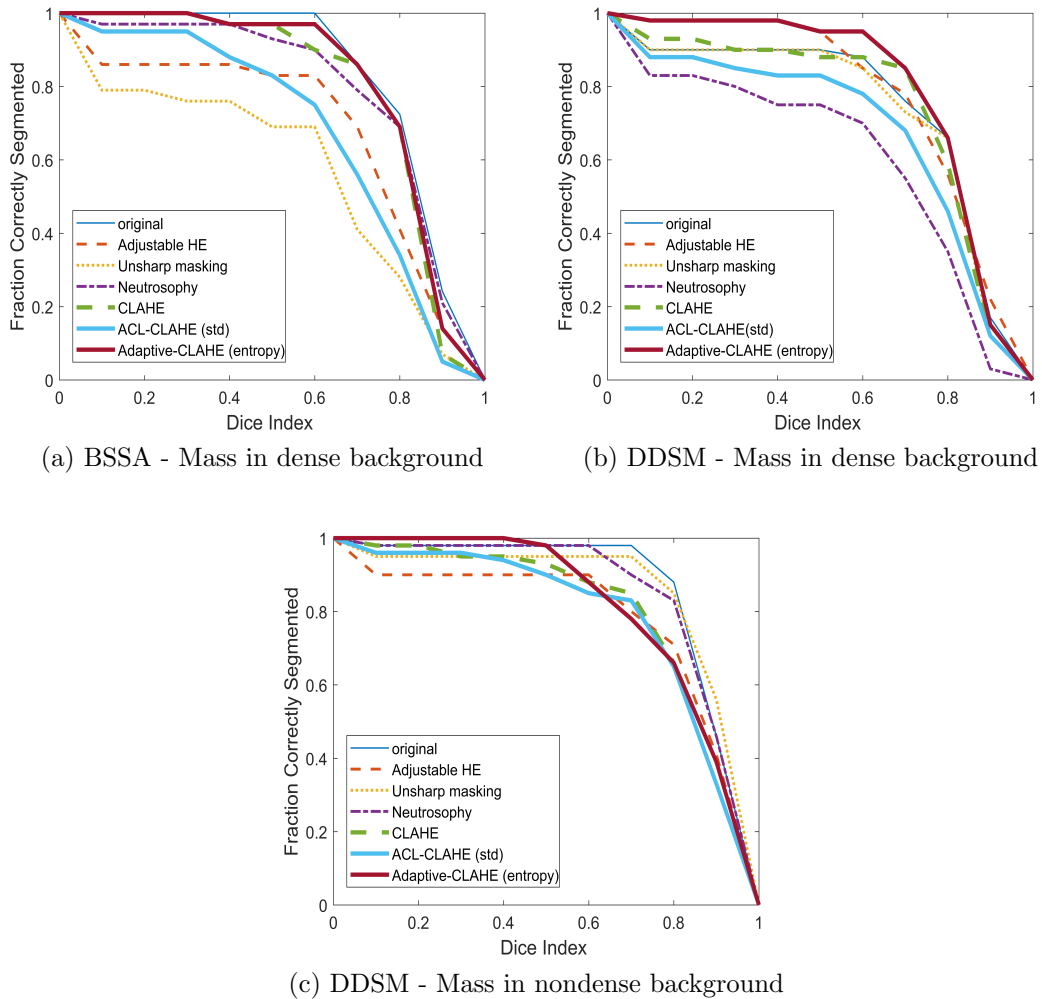
Figure 6.9: Performance comparison for SRM mass segmentation in local dense and nondense background with six different image enhancement methods and original image without any enhancement.

results shows that the proposed adaptive CLAHE with FCM outperformed all other techniques for both masses in dense background and masses in nondense background (see Table 6.1 and 6.2). In addition, the automatic parameter selection capability of the proposed adaptive CLAHE provides a superior advantage by removing manual effort and reducing accuracy issues.

The second method to improve mass segmentation was using optimized SRM. The entropy measure used to optimize the SRM segmentation algorithm was different from the one used for CLAHE enhancement optimization. For SRM, a component measure of entropy that calculates the layout entropy based on the number of SRM segmented regions and a region entropy that measures the uniformity within the regions were used. A different approach based on the union

Table 6.5: Performance comparison for SRM mass segmentation using Hausdorff Distance for BSSA (29 images) and the DDSM set for both mass in dense background (41 images) and nondense background (41 images) for each of the seven methods. Columns 2, 3 and 4 show the mean Hausdorff Distance for BSSA - mass in dense background, DDSM - mass in dense background and DDSM - mass in nondense background respectively. Standard deviation is shown in brackets.

Approach	BSSA - Mass in dense background	DDSM - Mass in dense background	DDSM - Mass in nondense background
Original (no enhancement)	3.79 (1.04)	3.70 (1.25)	**2.93 (1.17)**
Adjustable HE (Arici et al. 2009)	4.19 (0.88)	3.78 (0.91)	3.20 (1.09)
Unsharp Masking	4.53 (0.88)	3.78 (1.27)	2.99 (0.91)
Neutrosophic (Guo & Cheng 2009)	3.61 (1.09)	4.51 (0.92)	3.25 (0.95)
CLAHE	3.66 (1.07)	3.76 (1.11)	3.25 (1.30)
ACL-CLAHE (std) (Abbas et al. 2013)	**3.43 (0.79)**	3.86 (1.46)	3.49 (1.49)
Adaptive CLAHE (entropy)	3.48 (0.75)	**3.62 (0.98)**	3.10 (1.17)

of reasonable components inside the ground truth with Dice index as the performance measure was analyzed. In addition to Dice index, Hausdorff distance was also used as the performance measure. The experimental results show that the proposed adaptive SRM technique has the potential to produce effective mass segmentation in dense background while retaining high performance on local nondense background when merging the components (see Table 6.3). The result is particularly promising since without any enhancement, with the proposed automatic tuning, SRM was able to produce significantly better mass segmentation when compared to FCM results (Table 6.3). However, it has some limitations when compared to the first approach (adaptive CLAHE with FCM). The optimized SRM showed better performance with merging the segmented components. Finding a criteria to automatically merge the segmented components is not an easy task. In this chapter, ground truth was used to analyze the potential, it needs to be automatic when used in the mass localisation process. Due to time constraints, we were unable to find an efficient automatic merging criteria for SRM components and hence this is kept as future work arising from this thesis. Considering the above limitation of the proposed SRM technique, the proposed adaptive CLAHE with FCM will be used in the next mass localisation chapter for mass candidate generation.

Chapter 7

Automatic Mass Localisation

In Chapters 3, 4 and 5, we have presented different strategies using local features based on primitive micropatterns and their macro level representations to characterize breast mass tissues localized in dense background. In Chapter 6, we have presented two methods for improving breast mass segmentation in dense background. In this chapter, we present a framework for automatic mass localisation in dense breast using the techniques developed in the previous chapters. Section 7.1 describes the image dataset. Methodology adopted is described in Section 7.2. Experimental results obtained are presented in Section 7.3 followed by discussion and conclusion in Section 7.4.

7.1   Dataset

The data used in this chapter for experiments are the same as described in Section 6.1. Original image resolution was used to extract the texture features and downsampled images (by a factor of 8) were used for all the initial steps to save the processing time.

7.2   Methods

The proposed mass localisation approach consists of six major stages: breast region extraction using Otsu thresholding (Otsu 1979), contrast enhancement using the proposed adaptive CLAHE (Section 6.2.2), segmentation based on FCM clustering, selection of suspicious mass regions, feature extraction and classification.

7.2.1 Breast Region Extraction

Mammographic images normally contain irrelevant elements like labels, background etc., which can interfere with the accuracy of the mass CAD. Hence, the first step in the mass CAD is removing these unwanted elements. This step helps to focus the search for suspicious mass candidates to the breast area only and thus improve the mass localisation. Otsu thresholding method (Otsu 1979) was used to separate the foreground regions with high intensity from the dark background region. After thresholding the background intensity values are changed to zero. Morphological opening operation with a disk shaped structuring element with size 10 was applied to remove small objects from mammograms and filling was used to obtain a smooth breast border. Then the region with largest area was selected as the breast region. Figure 7.1 shows an example of breast region extraction.

(a)	(b)	(c)	(d)	(e)
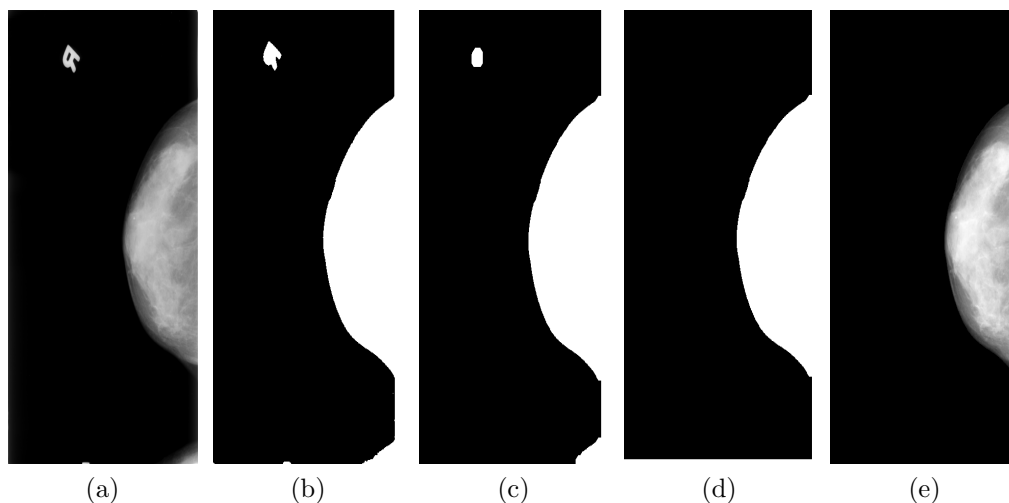
Figure 7.1: (a) Original mammogram (b) after Otsu thresholding (c) after morphological operation (d) largest area selection (e) mammogram after breast region extraction.

7.2.2 Contrast Enhancement using Adaptive CLAHE

As mentioned previously, one of the main difficulties in detecting masses in mammograms is similarity of brightness of the objects in the mammograms. Normal dense tissues and masses have similar X-ray attenuation and appear as bright regions in mammograms. This may increase the difficulty of mass localisation in dense breast. In order to alleviate this difficulty, contrast enhancement using adaptive CLAHE based on entropy, which was proposed in Chapter 6, is used.

The objective of this step is to enhance the contrast between the masses and their background in-order to improve the mass candidate segmentation and to improve the localisation accuracy. The adaptive CLAHE steps are explained in Chapter 6. Figure 7.2 shows an example of original mammogram (a) and enhanced mammogram (b).

<div align="center">(a)   (b)</div>
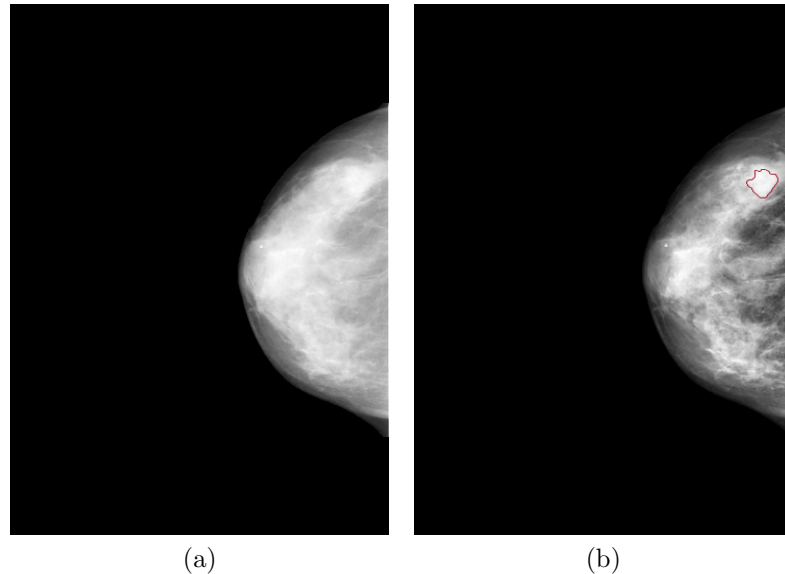
Figure 7.2: (a) Before contrast enhancement (b) after adaptive CLAHE contrast enhancement.

7.2.3 Breast Segmentation and Suspicious Mass Regions Selection

Adaptive CLAHE with FCM clustering, proposed in Chapter 6 was used to produce ROIs in this chapter. The number of clusters (10) for FCM was chosen based on the results from Chapter 6. The segmented mammograms were further processed by applying morphological filing of holes, followed by morphological opening operation, with the size 2 structuring element of disk type. The opening operation smooths contours, break narrow isthmuses and eliminates small islands. All connected components with area bigger than 50px were considered for further analysis.

In FCM, lower clusters have regions with low intensity values and top clusters have regions with high intensity values. For finding the suspicious mass regions, we considered only top clusters as mass regions have higher intensity compared to others. The top four clusters, in the case of DDSM, and the top three clusters,

in case of the BSSA, were considered. After removing the regions in the lower clusters of FCM resulted in 1905 ROIs for 41 mammograms (mass in dense background) collected from the DDSM database. In that, 75 were mass ROIs and 1830 were normal ROIs. Using core mass contours information, an automatically generated ROI was considered to be a true mass ROI only if it met two criteria: (1) the centroid of a segmented region was included in the annotated area and (2) more than 25% of the segmented region intersected with the true mass region (annotated region) (Choi & Ro 2012, Mudigonda et al. 2001, Dominguez & Nandi 2008, Eltonsy et al. 2007, Varela et al. 2007). The details of ROIs in each datset are provided in Table 7.1. None of the masses were missed in the process. Figure 7.3 shows an example of the mammogram with suspicious regions.

| (a) | (b) | (c) |
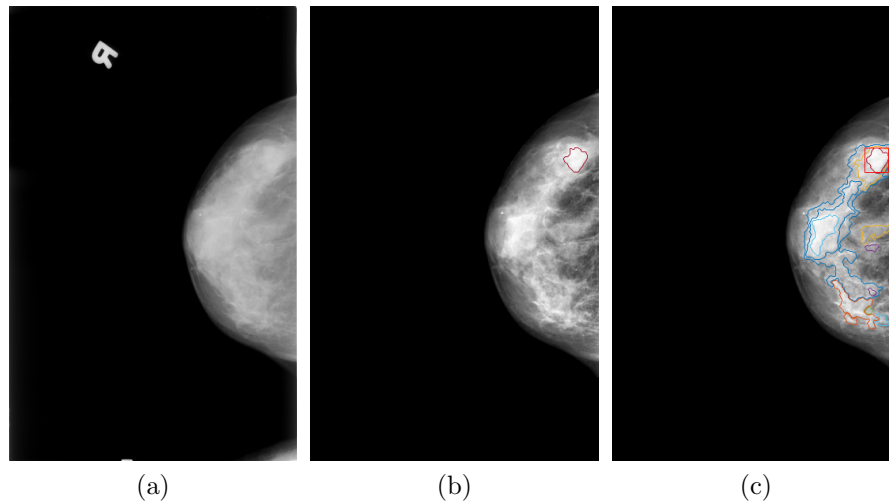
Figure 7.3: Example showing suspicious regions (a) original mammogram (b) after breast region extraction and contrast enhancement (c) suspicious regions with ground truth marked with a box.

Table 7.1: Number of ROIs extracted from each dataset. Dense refers to mass in dense background and nondense refers to mass in nondense background. The number underneath each dataset shows the number of mammograms in each set.

	BSSA (29)	DDSM - dense (41)	DDSM - nondense (41)
No. of mass regions	49	75	77
No.of normal regions	1708	1830	1851
ROIs/image	60.5	46.5	47

7.2.4 Feature Extraction

After suspicious regions identification, features were extracted to find the characteristics of the detected objects. The texture features that were proposed in this thesis (Chapters 3, 4 and 5) were extracted. The details of the feature extraction were given in the respective chapters.

In addition to the proposed texture features, some basic mass features (mean intensity, area and eccentricity) were also extracted to study the effect of combined feature set.

The texture feature extraction used the original resolution image. Although ROI contours were estimated initially at low resolution, they were expanded to original resolution prior to feature extraction. As mentioned in Section 3.4, using MATLAB, the smallest rectangular region fitting the contours were extracted and features were calculated from it. Red box shown in Figure 7.3 (c) is an example.

7.2.5 Classification

The best features found from each of the proposed texture analysis methods in previous chapters were included in the feature set for this study. As per the previous chapters, LDA and SVM-L were used for classification. As the data for classification was highly unbalanced (see Table 7.1), for training, minority class was over sampled using SMOTE (Synthetic Minority Over-Sampling Technique) (Chawla et al. 2002). Based on the features of the original dataset, this technique generates new instances of minority class from the nearest neighbors of line joining the minority class samples. In our study, SMOTE was used to generate same number of mass samples as normal ones to have a balanced dataset for training the classifier. Due to very small dataset, leave one out cross validation was used. For DDSM, the classifier was trained with ROIs from 40 mammograms and tested on ROIs from one mammograms and this was repeated 41 times. Similarly, for BSSA, the classifier was trained with ROIs from 28 mammograms and tested on ROIs from one mammograms and this was repeated 29 times.

7.3 Experimental Results

Table 7.2 shows the performance of features generated from individual approaches and their combined performance. The features selected in the previous chapters

(Chapters 3, 4 and 5) were analysed and the best results are shown. Results shown are for the features: *micropatterns* (Chapter 3) - $\{f_0, f_1, f_{255}\}$ at radius 1, 2 and 3, *macropatterns* (Chapter 4) - $\{f_0, f_7, f_{255}\}$ at levels 1 and 3, *superpixel pattern graph* (Chapter 5) - 3 graph features (average vertex degree, cluster coefficient, and giant connected component ratio) generated from scale 1 and 2 of $\{p_0, p_1, p_{255}\}$ combination. When the features from the three approaches were combined, all performance measures showed significant improvement for all three datasets.

Adding basic mass features (mean intensity, area and eccentricity) along with the proposed texture feature again improved the classification results. Table 7.3 show the combined results for LDA and SVM-L classifiers. Figure 7.4 shows the ROC curves for each of the classifiers (LDA, SVM-L) for BSSA, DDSM dense and nondense datasets. The proposed method achieved an AUC score up to 0.925 and 0.950 for mass in dense and nondense background respectively. The same features showed an AUC score of 0.923 when tested on BSSA dataset, with a model trained on DDSM dense dataset, and an AUC score of 0.907 was achieved when tested on DDSM - dense dataset, with a model trained on BSSA dataset. The results are summarized in Table 7.4.

(a) LDA   (b) SVM-L
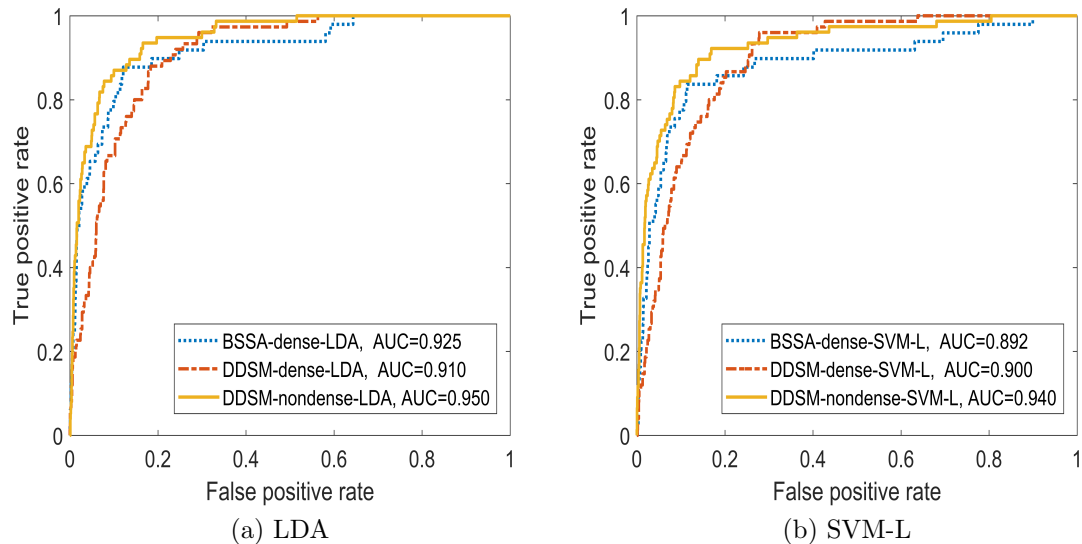
Figure 7.4: ROC curves illustrating the effectiveness of proposed approach for DDSM and BSSA datasets with LDA and SVM-L classifiers.

To evaluate the localisation performance of the CAD system based on proposed features (texture features combined with basic mass features), FROC

Table 7.2: Performance of individual feature extraction approaches (*micropatterns*, *macropatterns* and *superpixel patterns graph* developed in Chapters 3, 4 and 5 respectively) and their combined performance for BSSA, DDSM - dense and DDSM - nondense using LDA and SVM-L classifiers.

	BSSA - mass in dense background					
	LDA			SVM-L		
Features	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
micropatterns	0.875	0.861	0.742	0.880	0.860	0.810
macropatterns	0.834	0.825	0.721	0.820	0.840	0.720
superpixel patterns graph	0.802	0.800	0.710	0.810	0.800	0.710
combined - above three	0.901	0.880	0.802	0.913	0.880	0.801
	DDSM - mass in dense background					
	LDA			SVM-L		
Features	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
micropatterns	0.826	0.867	0.685	0.823	0.837	0.714
macropatterns	0.783	0.695	0.718	0.793	0.724	0.694
superpixel patterns graph	0.802	0.789	0.704	0.801	0.814	0.674
combined - above three	0.871	0.827	0.761	0.870	0.870	0.790
	DDSM - mass in nondense background					
	LDA			SVM-L		
Features	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
micropatterns	0.780	0.740	0.695	0.780	0.741	0.800
macropatterns	0.881	0.867	0.797	0.890	0.900	0.786
superpixel patterns graph	0.714	0.680	0.690	0.720	0.657	0.712
combined - above three	0.893	0.874	0.815	0.878	0.803	0.800

Table 7.3: Classification results for proposed texture features (*combined* Table 7.2) along with basic mass features.

	LDA			SVM-L		
Database	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
BSSA	0.925	0.877	0.852	0.892	0.780	0.870
DDSM - dense	0.910	0.883	0.816	0.900	0.760	0.850
DDSM - nondense	0.950	0.911	0.879	0.940	0.853	0.911

curves were used. Figure 7.5 show the FROC results obtained for LDA and SVM-L classifiers for three datasets (BSSA-dense, DDSM-dense and DDSM-nondense). For BSSA, we obtained a sensitivity of 80% at 2.7 false positive per image (FPI), while DDSM - dense obtained a sensitivity of 79% at 2.9 FPI and DDSM - non-

dense obtained a sensitivity of 80% at 2.8 FPI using LDA classifier. It should be noted that when compared to the initial suspicious ROIs selection, FPI fell from 58.8 (1708/29) to 2.7 (95% reduction from initial selection) with the proposed features for BSSA. For DDSM - dense, the reduction was from 44.6 (1830/41) to 2.9 (93% reduction) and for DDSM - nondense, from 45 (1851/41) to 2.8 (94% reduction). Both classification and localisation results indicate that the proposed features were equally good in dense and nondense background. Figure 7.6 shows the example of mammograms illustrating the effectiveness of the proposed solution for correctly finding the true mass from the false positive regions. From the figure, it is clear that the false positive regions are removed using the proposed approach without losing the true mass region.

Table 7.4: Classification results: DDSM used for training and BSSA used for testing (DDSM/BSSA), and BSSA used for training and DDSM used for testing (BSSA/DDSM). The combined feature set was used.

	DDSM/BSSA			BSSA/DDSM		
Classifiers	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
LDA	0.923	0.918	0.800	0.885	0.825	0.879
SVM-L	0.907	0.878	0.845	0.867	0.796	0.895

To our knowledge, there is no study in the literature that particularly focuses on dense breast, due to its challenging nature. A few studies have reported their mass localisation performance in dense breast from DDSM, even though their focus was not on dense breast. There is no evidence to show that all the masses used in these studies are located in dense background, even though they are of density 3 and 4. Tai et al. (2014) reported 80% sensitivity with 3 FPI in BI-RADS breast density 3 and 69% sensitivity with 4 FPI in breast density 4. Liu & Zeng (2015) reported 77% sensitivity with 4.9 FPI in BI-RADS breast density 3 and 57% sensitivity with 4.1 FPI in breast density 4. Compare to these results, our approach obtained a sensitivity of 79% at 2.9 FPI for DDSM mass in dense background. This shows that the features resulting from this study characterizes masses in dense background significantly better than the closest research available in literature.
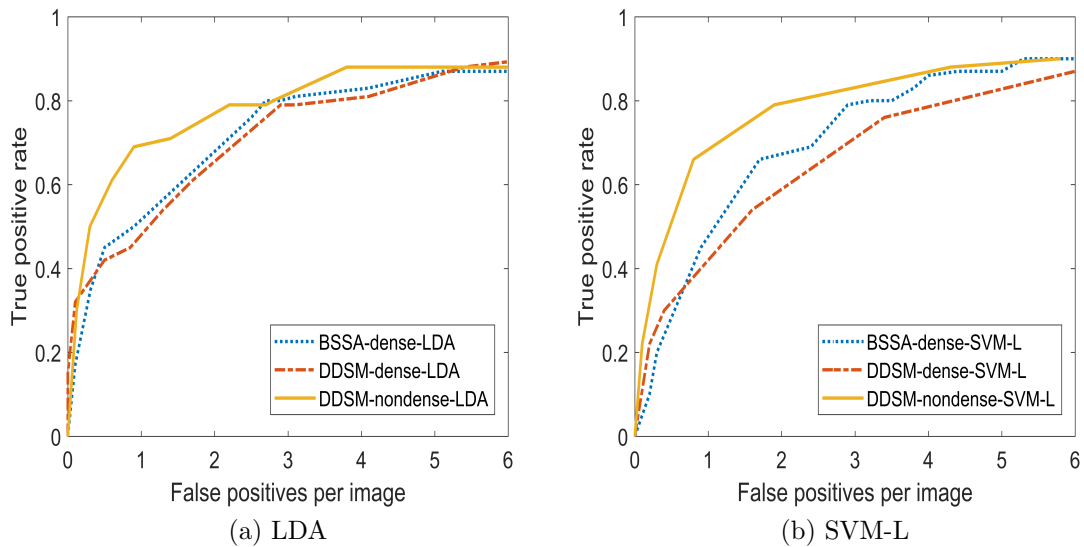
Figure 7.5: FROC curves illustrating the effectiveness of proposed approach for mass localisation for DDSM and BSSA datasets with LDA and SVM-L classifier.

7.4 Conclusion and Discussion

In this chapter, all the methods proposed in the previous chapters are evaluated using an automatic breast mass CAD system. Adaptive CLAHE was used for contrast enhancement and local geometrical features generated from micro and macro patterns were used to describe the mass texture features.

Performance of the proposed system was evaluated using ROC and FROC curves. The highest AUC scores of 0.925 and 0.910 were obtained for mass classification on BSSA and DDSM - dense respectively. To prove the worth of the proposed approach in the general case, experiment was conducted on DDSM masses in nondense background and results were comparable (DDSM - nondense: AUC = 0.950). For mass localisation, with the proposed automatic approach, we obtained a sensitivity of 80% with 2.7 FPI for BSSA. This gave a 95% FPI reduction, when compared with the initial number of false positives of 58.8. Similar results were achieved for DDSM - dense (sensitivity of 79% at 2.9 FPI) and DDSM - nondense (sensitivity of 80% at 2.8 FPI) datasets.

A significant reduction in false positives with high sensitivity shows the mass discrimination capability of the proposed features. It shows that it could be used in the clinical setting for mass localisation in dense breast, which is very challenging for radiologists. At present, radiologists are finding it very challenging
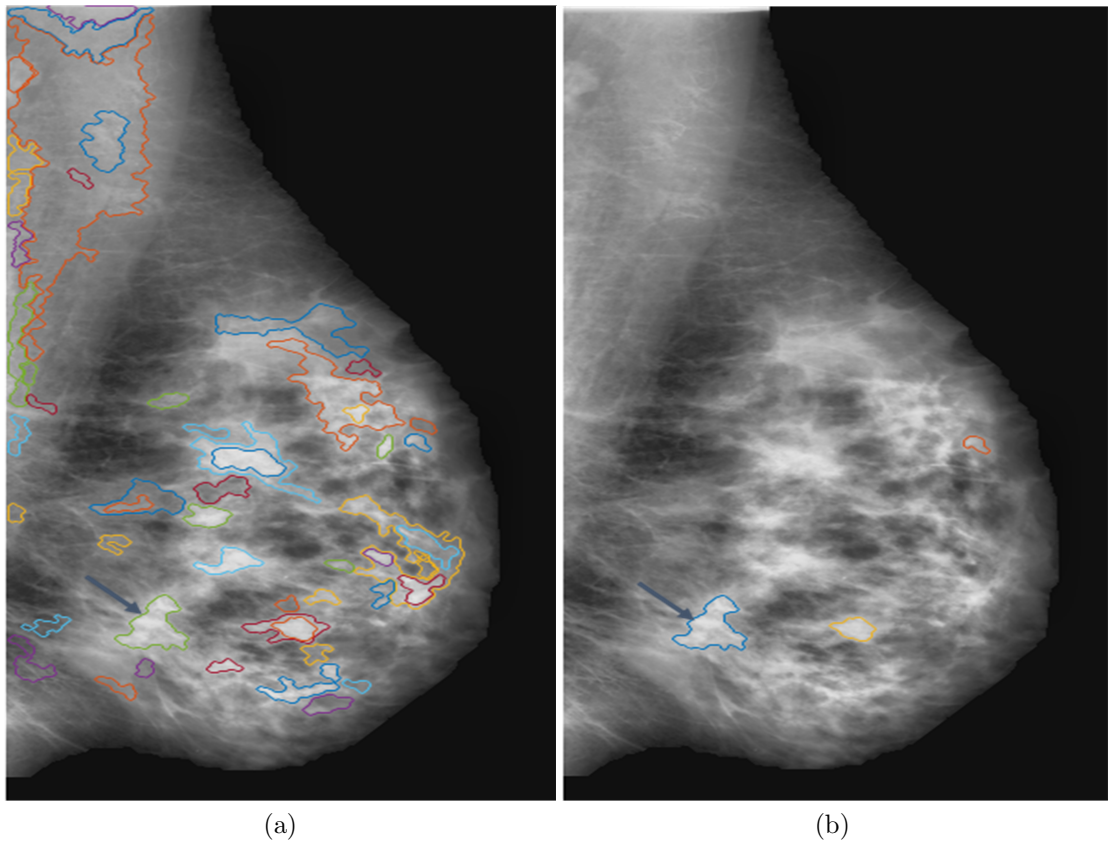
Figure 7.6: Example demonstrating the effectiveness of the proposed method. a) initial suspicious ROIs selection b) after classification using the proposed feature set. Mass region pointed at with an arrow.

to visually identify cancer in dense breast as normal dense breast tissue hides the tumor. This undetected, so untreated, masses/tumors could result in mortality. With the proposed CAD approach, highly suspicious regions are detected and, thus, allow radiologists to do a focused investigation on these regions. This could save many lives that otherwise may fall for undetected cancer.

Chapter 8

Conclusion and Future Work

8.1 Summary of the Thesis

This thesis presented a set of mammographic image analysis methods that can be incorporated into a CAD system for efficient mass localisation in dense breast. The thesis started with a brief introduction on breast cancer statistics, screening mammography, abnormalities in mammograms, problems with dense breast tissues and the usefulness of CAD in mammography. Following this, a literature review of texture analysis was presented. Then the main techniques used in a CAD system and related works in applications of texture analysis to mass detection/classification and mass detection in dense breast were described. From the literature review, it was found that mass detection in dense breast is a challenging task for radiologists and existing CAD. In addition, there was no evidence found in literature for structural texture analysis for mammographic masses. Hence, it is worthwhile developing effective structural texture analysis methods that can be incorporated into the CAD systems that can help radiologists in analysing masses that are localized in the dense background. Three different texture analysis methods were developed. Subsequently, two methods were developed for improving breast mass segmentation as a preprocessing step. Finally, all methods were incorporated into the automatic mass localisation system. The main contributions of this thesis to the literature are the texture analysis methods developed.

For dense breast mass texture analysis (Chapter 3, Chapter 4 and Chapter 5), which was the main aim of this thesis, three approaches, which can be viewed as an extension/improvement of one another, were developed. In the first approach, the characteristics of masses localized in dense background were analyzed using structured micropatterns generated using LBP technique at different radius lev-

els. These patterns can describe micro-structures like edges, lines, spots and flat areas, which refer back to Julesz Textons which are the basic elements of pre-attentive human perception (Julesz 1981). In addition, this approach provides a combination of structural and statistical texture analysis. Structural, as these patterns describes the micro structures in the image, and statistical, as proportion/density of these patterns form features in the study. The performance of the proposed approach was evaluated on ROIs (all localized in dense backgrounds of breasts) extracted from the DDSM and BSSA databases. The experimental results show that features generated from structured micropatterns can produce very effective and efficient texture descriptors of cancerous ROIs. The traditional histogram concatenation approach of extracting features from LBP is replaced by features based on combinatorial properties of the structured micropatterns. This allows for a huge reduction of the dimension in the feature space while keeping a high rate of mass classification. With few features (6 features) AUC scores of 0.961 for DDSM and 0.885 for BSSA were achieved.

Even though the pixel level approach showed high potential in mass discrimination, it lacks an ability to capture the macrostructures in the image and sensitive to noise. Hence, to overcome that, a new paradigm for texture analysis based on superpixel tessellation was proposed. In this paradigm, ROIs were first transformed into a superpixel tessellation using a regular grid approach and an irregular grid approach. The superpixel texture analysis paradigm has the potential to generate a range of new textures. In this thesis, from the superpixel tessellation, structured superpixel patterns were constructed to represent the macrostructures in the ROI image and density of each pattern formed the basis for mass classification. The performance of the proposed approach was evaluated on the same dataset used for the pixel level approach. The highest AUC scores of 0.961 on DDSM and 0.924 on BSSA were achieved with few features.

The features generated from the pixel based and superpixel based approaches captures the density/proportion of the structured patterns, but it did not reflect the spatial distribution and closeness of the patterns. In order to overcome this, spatial connectivity graphs on structured superpixel patterns were developed. The superpixel patterns connectivity graphs based approach is computationally efficient and provides a robust tool for characterizing and representing dense mass ROIs according to the spatial distribution of texture patterns. Four clinically recognizable features were extracted from the superpixel graphs and used for mass classification. The performance of the developed approaches were compared with related publications in the literature.

For breast mass segmentation (Chapter 6), two methods were developed. First one based on the combination of adaptive CLAHE and FCM and, second, an adaptive SRM. The performance of the developed methods was tested on BSSA and DDSM images: mammograms with masses in dense background. In addition, to investigate the performance in masses in nondense background, it was tested on mammograms with masses in nondense background taken from DDSM database. The experimental results showed a high percentage of acceptable mass segmentation for both methods for masses localized in dense and nondense background (adaptive CLAHE approach obtained 95% and 98% for masses in local dense and nondense background respectively). A direct comparison with four commonly used enhancement techniques from the literature and the original image was provided. The results demonstrate that the developed methods have a high potential to improve the breast mass segmentation in dense background while retaining high performance levels on local nondense background.

Finally, an overall evaluation of all the proposed methods described earlier was presented in Chapter 7, demonstrating the effectiveness of these methods for mammographic mass localisation. For masses in dense background, the proposed approach achieved a sensitivity of 79% with 2.9 FPI and a sensitivity of 80% with 2.7 FPI for DDSM and BSSA respectively. The experimental results shows that the proposed approaches can be incorporated in to a real clinical setting to help radiologists in identifying cancer localized in dense background. With the proposed CAD approach, highly suspicious regions are detected and thus allow radiologists to do a focused investigation on these regions. This could save many lives that otherwise may have been left undetected and, therefore untreated and result in mortality.

8.2   Future Work

Even though the proposed approaches in this thesis have showed high potential in enhancement, segmentation, classification and localisation of masses in dense regions of mammograms, there are some limitations and observations that are seeds for future work.

The optimized SRM technique (Chapter 6) showed a high potential for mass segmentation via merging of the SRM segmented components. In Chapter 6, ground truth was considered to select the components that needed to be merged. Due to limited time, finding a criteria to automatically merge the segmented SRM

components was kept as future work.

The methods used to generate superpixel tessellation and textures generated from the superpixel tessellation presented in this thesis are not the only ones (Chapters 4 and 5) possible. Other methods need to be explored in the future. In the graph based approach (Chapter 5), instead of topological features, graph similarity/dissimilarity measures between cancer and normal superpixel pattern graphs could be explored in the future.

In mass localisation approach (Chapter 7), pectoral muscle removal from mammograms could be included in the preprocessing stage to further reduce false positives. In addition, different mass segmentation algorithms other than FCM and SRM could be analysed for their potential.

The proposed approaches need to be applied to different computer vision applications such as cancer detection in histopathological images especially the graph study, which has high potential for cell graphs. The texture approaches could also be applied to other imaging like CT and MRI and cancer like abnormality detection in lung and brain.

The availability of dense breast data is a challenge that, if we could overcome in the future, can lead to efficient learning.

8.3 Contributions of this thesis

The main contributions of this thesis are summarized below:

1. A novel approach of extracting structured micropatterns using LBP technique for dense mass ROI classification was developed. This technique was tested with images from DDSM and BSSA. The proposed approach produced comparative results compared to state of the art work with huge reductions in the feature dimensionality.

2. A new paradigm for texture analysis based on superpixel tessellation was developed. From the superpixel tessellation, structured superpixel patterns were constructed to represent the macro-structures in the ROI. The density of each pattern formed the basis for mass classification.

3. A new graph representation for dense mass classification was developed. A set of structured superpixel patterns graphs were constructed to analyse the

characteristics of masses. The topology of superpixel patterns, captured by using spatial connectivity graphs, revealed significant differences between cancerous and healthy areas of breasts.

4. A self-adjusted mammogram contrast enhancement solution called adaptive CLAHE was developed. The method automatically tunes the clip limit and block size of the CLAHE algorithm. The proposed contrast enhancement algorithm improved the subsequent mass segmentation process.

5. Finally, an automated breast mass localisation system was developed that has the potential to localise masses located in dense background.

References

Abbas, Q., Celebi, M. E. & Garcia, I. F. (2013), 'Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system', *Biomedical Signal Processing and Control* .

Abdel-Nasser, M., Rashwan, H. A., Puig, D. et al. (2015), 'Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern', *Expert Systems with Applications* **42**(24), 9499 – 9511.

Achanta, R., Shaji, A., Smith, K. et al. (2012), 'SLIC superpixels compared to state-of-the-art superpixel methods', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11), 2274–2282.

ACR (2003), 'Breast Imaging Reporting and Data System (BI-RADS)', *4th edn. American College of Radiology, Reston* .

ACR (2014), 'ACR BI-RADS Atlas Fifth Edition', `https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/BIRADS-Reference-Card.pdf`. (Accessed: 17 April 2018).

ACR (2018), 'Mammography facts', `https://www.mammographysaveslives.org/Facts`. (Accessed: 20 August 2018).

ACS (2018), 'Cancer facts and figures 2018', `https://https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf`. (Accessed: 8 February 2019).

ACSBC (2018), 'American Cancer Society Guidelines for the Early Detection of Cancer', `https://www.cancer.org/healthy/find-cancer-early/cancer-screening-guidelines/american-cancer-society-guidelines-for-the-early-detection-of-cancer.html`. (Accessed: 8 February 2019).

Ahuja, N. (1982), 'Dot pattern processing using voronoi neighborhoods', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-4**(3), 336–343.

AIHW (2017), 'Australian Institute of Health and Welfare (AIHW) 2017 Australian Cancer Incidence and Mortality (ACIM) books', `http://www.aihw.gov.au/acim-books`. (Accessed: 20 January 2018).

AIHWC (2017), 'Australian Institute of Health and Welfare Canberra, Cat. no. CAN 100', `https://www.aihw.gov.au/getmedia/3da1f3c2-30f0-4475-8aed-1f19f8e16d48/20066-cancer-2017.pdf.aspx?inline=true`. (Accessed: 20 January 2018).

Amroabadi, S. H., Ahmadzadeh, M. R. & Hekmatnia, A. (2011), Mass detection in mammograms using GA based PCA and Haralick features selection, *in* 'Electrical Engineering (ICEE), 2011 19th Iranian Conference on', IEEE, Tehran, Iran, pp. 1–4.

Arici, T., Dikbas, S. & Altunbasak, Y. (2009), 'A histogram modification framework and its application for image contrast enhancement', *IEEE Transactions on Image Processing* **18**(9), 1921–1935.

Azavedo, E., Zackrisson, S. & others, I. M. (2012), 'Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review', *BMC medical imaging* **12**(1), 22.

Bai, Z. & Hatzinakos, D. (2010), LBP-based biometric hashing scheme for human authentication, *in* '2010 11th International Conference on Control Automation Robotics Vision', pp. 1842–1847.

Bajger, M., Ma, F. & Bottema, M. J. (2009), Automatic tuning of MST segmentation of mammograms for registration and mass detection algorithms, *in* 'Digital Image Computing: Techniques and Applications (DICTA2009)', pp. 400–407.

Barbara, W., van de Wouwer Gert, Samir, K. et al. (1999), 'Computer assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis', *Cytometry* **35**(1), 23–29.

Bilgin, C. C., Bullough, P., Plopper, G. E. et al. (2010), 'ECM-aware cell-graph mining for bone tissue modeling and classification', *Data mining and knowledge discovery* **20**(3), 416–438.

Bilgin, C., Demir, C., Nagi, C. et al. (2007), Cell-graph mining for breast tissue modeling and classification, *in* 'Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE', pp. 5311–5314.

Birdwell, R. L., Bandokar, P. & Ikeda, D. M. (2005), 'Computer aided detection with screening mammography in a university hospital setting', *Radiology* **236**(2), 451–457.

Blostein, D. & Ahuja, N. (1989), 'Shape from texture: Integrating texture-element extraction and surface estimation', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(12), 1233–1251.

Blot, L. & Zwiggelaar, R. (2005), 'A volumetric approach to glandularity estimation in mammography: a feasibility study', *Physics in Medicine and Biology* **50**(4), 695–708.

Boyd, N. F., Byng, J. W., Jong, R. A. et al. (1995), 'Quantitative classification of mammographic densities and breast cancer risk: Results from the Canadian National Breast Screening Study.', *J Natl Cancer Inst* **87**(9), 670–675.

Boyd, N. F., Dite, G. S., Stone, J. et al. (2002), 'Heritability of mammographic density: a risk factor for breast cancer', *New England Journal of Medicine* **347**(12), 886–894.

Boyd, N. F., Guo, H., J.Martin, L. et al. (2007), 'Mammographic density and the risk and detection of breast cancer', *New England Journal of Medicine* **356**(3), 227–236.

Bradley, A. P. (1997), 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognition* **30**(7), 1145–1159.

Brem, R. F., Baum, J., Lechner, M. et al. (2003), 'Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial', *American Journal of Roentgenology* **181**(3), 687–693.

Brem, R. F., Hoffmeister, J. W., Rapelyea, J. A. et al. (2005), 'Impact of breast density on computer-aided detection for breast cancer', *American Journal of Roentgenology* **184**(2), 439–444.

Brzakovic, D., Luo, X. M. & Brzakovic, P. (1990), 'An approach to automated detection of tumors in mammograms', *IEEE Transactions on Medical Imaging* **9**(3), 233–241.

BSA (2019), 'Breast Screen Australia - Policy and Program', `http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/policy`. (Accessed: 8 February 2019).

Buciu, I. & Gacsadi, A. (2011), 'Directional features for automatic tumor classification of mammogram images', *Biomedical Signal Processing and Control* **6**(4), 370–378.

Buist, D. S. M., Porter, P. L., Lehman, C. et al. (2004), 'Factors contributing to mammography failure in women aged 40-49 years', *Journal of the National Cancer Institute* **96**(19), 1432–1440.

Byrne, C., Schairer, C., Wolfe, J. et al. (1995), 'Mammographic features and breast cancer risk: effects with time, age, and menopause status', *JNCI: Journal of the National Cancer Institute* **87**(21), 1622–1629.

Castellano, C. R., Nunez, C. V., Boy, R. C. et al. (2011), 'Impact of mammographic breast density on computer-assisted detection (CAD) in a breast imaging department', *Radiologia* **53**(5), 456–461.

Chakraborty, J., Midya, A., Mukhopadhyay, S. et al. (2013), Automatic characterization of masses in mammograms, *in* '2013 6th International Conference on Biomedical Engineering and Informatics', pp. 111–115.

Chakraborty, J., Midya, A. & Rabidas, R. (2018), 'Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns', *Expert Systems with Applications* **99**(1), 168–179.

Chan, H. P., Wei, D., Helvie, M. A. et al. (1995), 'Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space', *Phys. Med. Biol.* **40**(5), 857–876.

Chawla, N. V., Bowyer, K. W., Hall, L. O. et al. (2002), 'SMOTE: synthetic minority over-sampling technique', *Journal of artificial intelligence research* **16**(2002), 321–357.

Chen, Z. (2013), Mammographic Image Analysis: Risk Assessment and Microcalcification Classifcation Aspects, PhD thesis, Aberystwyth University, United Kingdom.

Chen, Z., Strange, H., Oliver, A. et al. (2015), 'Topological modeling and classification of mammographic microcalcification clusters', *IEEE Transactions on Biomedical Engineering* **62**(4), 1203–1214.

Cheng, E., Xie, N., Ling, H. et al. (2010), Mammographic image classification using histogram intersection, *in* '2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro', pp. 197–200.

Cheng, H., Cai, X., Chen, X. et al. (2003), 'Computer-aided detection and classification of microcalcifications in mammograms: a survey', *Pattern Recognition* **36**(12), 2967–2991.

Cheng, H. D., Shi, X. J., Min, R. et al. (2006), 'Approaches for automated detection and classification of masses in mammograms', *Pattern Recognition* **39**(4), 646–668.

Choi, J. & Ro, Y. M. (2012), 'Multiresolution local binary pattern texture analysis combined with variable selection for application to false-positive reduction in computer-aided detection of breast masses on mammograms', *Physics in Medicine & Biology* **57**(21), 7029–7052.

Choi, J. Y., Kim, D. H., Plataniotis, K. N. et al. (2016), 'Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography', *Expert Systems with Applications* **46**, 106–121.

Chokri, F. & Farida, M. H. (2017), 'Mammographic mass classification according to Bi-RADS lexicon', *IET Computer Vision* **11**(3), 189–198.

Christoyianni, I., Dermatas, E. & Kokkinakis, G. (2000), 'Fast detection of masses in computer-aided mammography', *IEEE Signal Process Mag.* **17**(1), 54–64.

Chuand, J., Min, H., Liu, L. et al. (2015), 'A novel computer aided breast mass detection scheme based on morphological enhancement and SLIC superpixel segmentation', *Medical physics* **42**(7), 3859–3869.

CRUK (2015), 'Cancer Research UK, Breast cancer statistics', `http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero`. (Accessed: 16 April 2018).

Cupples, T. E., Cunningham, J. E. & Reynolds, J. C. (2005), 'Impact of computer-aided detection in a regional screening mammography program', *American Journal of Roentgenology* **185**(4), 944–950.

da Rocha, S. V., Junior, G. B., Silva, A. C. et al. (2016), 'Texture analysis of masses malignant in mammograms images using a combined approach of diversity index and local binary patterns distribution', *Expert Systems with Applications* **66**, 7–19.

de Oliveira Silva, L. C., Barros, A. K. et al. (2017), 'Detecting masses in dense breast using independent component analysis', *Artificial intelligence in medicine* **80**, 29–38.

Dhahbi, S., Barhoumi, W., Kurek, J. et al. (2018), 'False-positive reduction in computer-aided mass detection using mammographic texture analysis and classification', *Computer methods and programs in biomedicine* **160**, 75–83.

Dheeba, J., Singh, N. A. & Selvi, S. T. (2014), 'Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach', *Journal of Biomedical Informatics* **49**, 45–52.

Dhungel, N., Carneiro, G. & Bradley, A. P. (2016), The automated learning of deep features for breast mass classification from mammograms, *in* 'International Conference on Medical Image Computing and Computer-Assisted Intervention', Springer, pp. 106–114.

Dhungel, N., Carneiro, G. & Bradley, A. P. (2017), 'A deep learning approach for the analysis of masses in mammograms with minimal user intervention', *Medical Image Analysis* **37**, 114 – 128.

Dice, L. R. (1945), 'Measures of the amount of ecologic association between species', *Ecology* **26**(3), 297—-302.

Dinnes, J., Moss, S., Melia, J. et al. (2001), 'Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review', *The Breast* **10**(6), 455–463.

Dominguez, A. R. & Nandi, A. K. (2008), 'Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection', *Computerized Medical Imaging and Graphics* **32**(4), 304 – 315.

Don, S., Choi, E. & Min, D. (2011), Breast mass segmentation in digital mammography using graph cuts, *in* 'Convergence and Hybrid Information Technology', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 88–96.

Drucker, F. & MacCormick, J. (2009), Fast superpixels for video analysis, *in* '2009 Workshop on Motion and Video Computing (WMVC)', pp. 1–8.

Eadie, L. H., Taylor, P. & Gibson, A. P. (2012), 'A systematic review of computer-assisted diagnosis in diagnostic cancer imaging', *European journal of radiology* **81**(1), e70–e76.

Eltonsy, N. H., Tourassi, G. D. & Elmaghraby, A. S. (2007), 'A concentric morphology model for the detection of masses in mammography', *IEEE Transactions on Medical Imaging* **26**(6), 880–889.

Endres, I. & Hoiem, D. (2014), 'Category-independent object proposals with diverse ranking', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(2), 222–234.

Ferlay, J., Colombet, M., Soerjomataram, I. et al. (2018), 'Cancer incidence and mortality patterns in europe: Estimates for 40 countries and 25 major cancers in 2018', *European Journal of Cancer* .

Fisher, R. A. (1936), 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics* **7**(2), 179–188.

Freer, T. W. & Ulissey, M. J. (2001), 'Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center', *Radiology* **220**(3), 781–786.

Freixenet, J., Oliver, A., Martí, R. et al. (2008), 'Eigendetection of masses considering false positive reduction and breast density information', *Medical physics* **35**(5), 1840–1853.

Friedman, J. H. (1997), 'On bias, variance, 0/1—loss, and the curse-of-dimensionality', *Data Mining and Knowledge Discovery* **1**(1), 55–77.

Friedman, J., Hastie, T. & Tibshirani, R. (2001), *The elements of statistical learning*, Vol. 1, Springer series in statistics New York.

Galloway, M. M. (1975), 'Texture analysis using gray level run lengths', *Computer Graphics and Image Processing* **4**(2), 172–179.

Ganesan, K., Acharya, R. U., Chua, C. K. et al. (2013*a*), 'Decision support system for breast cancer detection using mammograms', *J Engineering in Medicine* **227**(7), 721–732.

Ganesan, K., Acharya, U. R., Chua, C. K. et al. (2013*b*), 'Computer-Aided Breast Cancer Detection Using Mammograms: A Review', *IEEE reviews in Biomedical Engineering* **6**, 77–98.

Gargouri, N., Masmoudi, A. D. & Abid, D. S. M. R. (2012), 'A New GLLD Operator for Mass Detection in Digital Mammograms ', *International Journal of Biomedical Imaging* **2012**, 1–13.

George, M., Denton, E. & Zwiggelaar, R. (2018), Mammogram breast density classification using mean-elliptical local binary patterns, *in* '14th International Workshop on Breast Imaging (IWBI 2018)', Vol. 10718, International Society for Optics and Photonics, p. 107180B.

Gilbert, F. J., Astley, S. M., Gillan, M. G. et al. (2008), 'Single reading with computer-aided detection for screening mammography', *New England Journal of Medicine* **359**(16), 1675–1684.

Giordano, D. L., Karsa, L. V., Tomatis, M. et al. (2012), 'Mammographic screening programmes in europe: Organization, coverage and participation', *Journal of Medical Screening* **19**(1_suppl), 72–82.

Gonzalez, R. C. & Woods, R. E. (1992), *Digital Image Processing*, Addison-Wesley.

Gram, I. T., Funkhouser, E. B. & Tabar, L. (1997), 'The Tabar classification of mammographic parenchymal patterns', *European Journal of Radiology* **24**(2), 131–136.

Gunduz, C., Yener, B. & Gultekin, S. H. (2004), 'The cell graphs of cancer', *Bioinformatics* **20**(suppl_1), i145–i151.

Guo, Y. & Cheng, H. D. (2009), 'New neutrosophic approach to image segmentation', *Pattern Recognition* **42**(5), 587–595.

Gur, D. (2007), 'Digital mammography: Do we need to convert now?', *Radiology* **245**(1), 10–11.

Guyon, I., Boser, B. & Vapnik, V. (1993), Automatic capacity tuning of very large vc-dimension classifiers, *in* 'Advances in Neural Information Processing Systems', Morgan Kaufmann, pp. 147–155.

Hadjidemetriou, E., Grossberg, M. & Nayar, S. (2004), 'Multiresolution histograms and their use for recognition', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **26**(7), 831–847.

Hafiane, A., Seetharaman, G., Palaniappan, K. & Zavidovique, B. (2008), *Rotationally Invariant Hashing of Median Binary Patterns for Texture Classification*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 619–629.

Hao, Z., Wang, Q., Ren, H. et al. (2012), Multiscale superpixel classification for tumor segmentation in breast ultrasound images, *in* '2012 19th IEEE International Conference on Image Processing', pp. 2817–2820.

Haralick, R. M. (1979), 'Statistical and structural approaches to texture', *Proceedings of the IEEE* **67**(5), 786–804.

Haralick, R. M., Shanmugam, K. & Dinstein, I. (1973), 'Textural features for image classification', *IEEE Trans, Systems, Man, and Cybernetics SMC* **3**(6), 610–621.

Harvey, J. A. & Bovbjerg, V. E. (2004), 'Quantitative assessment of mammographic breast density: relationship with breast cancer risk', *Radiology* **230**(1), 29–41.

Harvey, S. C., Geller, B., Oppenheimer, R. G. et al. (2003), 'Increase in cancer detection and recall rates with independent double interpretation of screening mammography', *American Journal of Roentgenology* **180**(5), 1461–1467.

Hastie, T. J., Tibshirani, R. J. & Friedman, J. H. (2009), *The elements of statistical learning : data mining, inference, and prediction*, Springer series in statistics.

He, D.-C. & Wang, L. (1990), 'Texture unit, texture spectrum, and texture analysis', *Geoscience and Remote Sensing, IEEE Transactions on* **28**(4), 509–512.

He, W., Denton, E. R., Stafford, K. et al. (2011), 'Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments', *Biomedical Signal Processing and Control* **6**(3), 321–329.

Heath, M., Bowyer, K., Kopans, D. et al. (2001), The Digital Database for Screening Mammography, *in* 'Proceedings of the Fifth International Workshop on Digital Mammography', Medical Physics Publishing, pp. 212–218.

Heikkila, M. & Pietikainen, M. (2006), 'A texture-based method for modeling the background and detecting moving objects', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(4), 657–662.

Helvie, M. (2007), 'Improving mammographic interpretation: double reading and computer-aided diagnosis', *Radiologic Clinics of North America* **45**(5), 801–811.

Hild, M. & Shirai, Y. (1993), 'Extraction of texture elements from images of shaded scenes', *Pattern Recognition* **26**(8), 1177 – 1191.

Ho, T. K. (1995), Random decision forests, *in* 'Proceedings of 3rd international conference on document analysis and recognition', Vol. 1, IEEE, pp. 278–282.

Ho, W. T. & Lam, P. W. (2003), 'Clinical performance of computer-assisted detection (CAD) in detecting carcinoma in breasts of different densities', *Radiology* **58**(2), 133–136.

Hou, N., Hong, S., Wang, W. et al. (2013), 'Hormone replacement therapy and breast cancer: heterogeneous risks by race, weight, and breast density', *J Natl Cancer Inst* **105**(18), 1365–1372.

Hussain, M. (2013), 'False-positive reduction in mammography using multiscale spatial weber law descriptor and support vector machines', *Neural Computing and Applications* **25**(1), 83–93.

Hussain, M., Khan, S., Muhammad, G. et al. (2014), 'Effective Extraction of Gabor Features for False Positive Reduction and Mass Classification in Mammography', *Applied Mathematics and Information Sciences* **8**(1L), 397–412.

Jo, E. B., Lee, J. H., Park, J. Y. & Kim, S. M. (2013), Detection of Breast Cancer Based on Texture Analysis from Digital Mammograms, *in* 'Intelligent Autonomous Systems 12', Vol. 2 of *Advances in Intelligent Systems and Computing*, Springer, pp. 893–900.

Julesz, B. (1981), 'Textons, the elements of texture perception, and their interactions.', *Nature* **290**(5802), 91–97.

Junior, G. B., de Paiva, A. C., Silva, A. C. et al. (2009), 'Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and {SVM} ', *Computers in Biology and Medicine* **39**(12), 1063 – 1072.

Kallergi, M., Woods, K., Clarke, L. P. et al. (1992), 'Image segmentation in digital mammography: comparison of local thresholding and region growing algorithms', *Computerized Medical Imaging and Graphics* **16**(5), 231–323.

Kanadam, K. P. & Chereddy, S. R. (2016), 'Mammogram classification using sparse-ROI: A novel representation to arbitrary shaped masses', *Expert Systems with Applications* **57**, 204–213.

Kapur, J., Sahoo, P. & Wong, A. (1985), 'A new method for gray-level picture thresholding using the entropy of the histogram', *Computer Vision, Graphics, and Image Processing* **29**(3), 273 – 285.

Ke, L., Mu, N. & Kang, Y. (2010), Mass computer-aided diagnosis method in mammogram based on texture features, *in* '2010 3rd International Conference on Biomedical Engineering and Informatics', Vol. 1, pp. 354–357.

Keramidas, E. G., Iakovidis, D. K., Maroulis, D. et al. (2008), Thyroid texture representation via noise resistant image features, *in* '21st IEEE International Symposium on Computer-Based Medical Systems, 2008', pp. 560–565.

Khuzi, A. M., Besar, R., Zaki, W. W. et al. (2009), 'Identification of masses in digital mammogram using gray level co-occurrence matrices', *Biomedical imaging and intervention journal* **5**(3), e17.

Kolb, T., Lichy, J. & Newhouse, J. (2002), 'Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of the factors that influence them: an analysis of 27,825 patient evaluations', *Radiology* **225**(1), 165–175.

Kooi, T., Litjens, G., van Ginneken, B. et al. (2017), 'Large scale deep learning for computer aided detection of mammographic lesions', *Medical Image Analysis* **35**, 303 – 312.

Laws, K. I. (1980), Textured image segmentation, Technical report, University of Southern California Los Angeles Image Processing INST.

Lee, H. & Kim, J. (2016), Segmentation of overlapping cervical cells in microscopic images with superpixel partitioning and cell-wise contour refinement, *in* 'Computer Vision and Pattern Recognition (CVPR) Workshop on Computer Vision for Microscopy Image Analysis (CVMI)', pp. 63–69.

Levy, D. & Jain, A. (2016), 'Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks', *CoRR* .

Li, H., Wang, Y., Liu, K. R. et al. (2001), 'Computerized radiographic mass detection. i. lesion site selection by morphological enhancement and contextual segmentation', *IEEE Transactions on Medical Imaging* **20**(4), 289–301.

Li, X.-Z., Williams, S. & Bottema, M. J. (2014), 'Texture and region dependent breast cancer risk assessment from screening mammograms', *Pattern Recognition Letters* **36**, 117 – 124.

Liao, C.-J. & Chen, S.-Y. (2002), 'Complementary retrieval for distorted images', *Pattern Recognition* **35**(8), 1705 – 1722.

Liao, S., Zhu, X., Lei, Z. et al. (2007), *Learning Multi-scale Block Local Binary Patterns for Face Recognition*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 828–837.

Lin, G. S., Chang, Y. C., Yeh, W. C. et al. (2014), 'A pyramid-based mass detection scheme based on texture analysis and neutral classification for digital mammograms', *Biomedical Engineering: Applications, Basis and Communications* **26**(4), 1–9.

Liu, L., Lao, S., Fieguth, P. W. et al. (2016), 'Median robust extended local binary pattern for texture classification', *IEEE Transactions on Image Processing* **25**(3), 1368–1381.

Liu, X., Xu, X., Liu, J. et al. (2011), A new automatic method for mass detection in mammography with false positives reduction by supported vector machine, *in* '2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)', Vol. 1, Shanghai, China, pp. 33–37.

Liu, X. & Zeng, Z. (2015), 'A new automatic mass detection method for breast cancer with false positive reduction', *Neurocomputing* **152**, 388 – 402.

Llado, X., Oliver, A., Freixenet, J. et al. (2009), 'A textural approach for mass false positive reduction in mammography', *Comput. Med. Imaging Graph.* **33**(6), 415–422.

Lowe, D. (1999), Object recognition from local scale-invariant features, *in* ' Proceedings of the Seventh IEEE International Conference on Computer Vision', Vol. 2, pp. 1150–1157.

Lu, S. & Fu, K. (1978), 'A syntactic approach to texture analysis', *Computer Graphics and Image Processing* **7**(3), 303 – 330.

Ma, F., Bajger, M. & Bottema, M. J. (2008), A graph matching based automatic regional registration method for sequential mammogram analysis, *in* 'Proc.SPIE', Vol. 6915, pp. 6915 – 11.

Ma, F., Bajger, M. & Bottema, M. J. (2009), Automatic mass segmentation based on adaptive pyramid and sublevel set analysis, *in* '2009 Digital Image Computing: Techniques and Applications', pp. 236–241.

Ma, F., Bajger, M., Slavotinek, J. P. et al. (2007), 'Two graph theory based methods for identifying the pectoral muscle in mammograms', *Pattern Recognition* **40**(9), 2592 – 2602.

MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, *in* 'Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics', University of California Press, Berkeley, Calif., pp. 281–297.

Maggio, C. D. (2004), 'State of the art of current modalities for the diagnosis of breast lesions', *European Journal of Nuclear Medicine and Molecular Imaging* **31**(1), 56–69.

Maitra, I. K., Nag, S. & Bandyopadhyay, S. K. (2012), 'Technique for preprocessing of digital mammogram', *Computer Methods and Programs in Biomedicine* **107**(2), 175—-188.

Malich, A., Fischer, D., Facius, M. et al. (2006), 'Effect of breast density on computer aided detection', *Journal of Digital Imaging* **18**(3), 227–233.

Mandelbrot, B. (1977), *Fractals*, Freeman San Francisco.

Mandelson, M. T., Oestreicher, N., Porter, P. L. et al. (2000), 'Breast density as a predictor of mammographic detection: Comparison of interval and screen detected cancers', *Journal of the National Cancer Institute* **92**(13), 1081–1087.

Manso, A. G., Orellana, C. J. G., Velasco, H. G. et al. (2013), 'Study of the effect of breast tissue density on detection of masses in mammograms', *Computational and Mathematical Methods in Medicine* **2013**(213794), 10.

McCormack, V. A. & Silva, D. S. I. (2006), 'Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis', *Cancer Epidemiol Biomarkers* **15**(6), 1159–1169.

McSweeney, M. B., Sprawls, P. & Egan, R. L. (1983), *Early Detection of Breast Cancer*, Springer.

Mencattini, A., Salmeri, M., Lojacono, R. et al. (2008), 'Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing', *IEEE transactions on instrumentation and measurement* **57**(7), 1422–1430.

Midya, A. & Chakraborty, J. (2015), Classification of benign and malignant masses in mammograms using multi-resolution analysis of oriented patterns, *in* '12th IEEE International Symposium on Biomedical Imaging, ISBI 2015, Brooklyn, NY, USA, April 16-19, 2015', pp. 411–414.

Miller, P. & Astley, S. (1992), 'Classification of breast tissue by texture analysis', *Image and Vision Computing* **10**(5), 277–282.

Min, B. S., Lim, D. K., Kim, S. J. et al. (2013), 'A Novel Method of Determining Parameters of CLAHE Based on Image Entropy', *International Journal of Software Engineering and Its Applications* **7**(5), 113–120.

Mirmehdi, M. (2008), *Handbook of texture analysis*, Imperial College Press.

Moreira, I. C., Amaral, I., Domingues, I. et al. (2012), 'Inbreast: toward a full-field digital mammographic database', *Academic radiology* **19**(2), 236–248.

Morrow, W. M., Paranjape, R. B. & Rangayyan, R. M. (1992), 'Region-based contrast enhancement of mammograms', *IEEE Transactions on Medical Imaging* **11**(3), 392–406.

Morton, M. J., Whaley, D. H., Brandt, K. R. et al. (2006), 'Screening mammograms: Interpretation with computer-aided detection prospective evaluation', *Radiology* **239**(2), 375–383.

Mudigonda, N. R., Rangayyan, R. M. & Desaultels, J. E. L. (2001), 'Detection of breast masses in mammograms by density slicing and texture flow-field analysis', *IEEE Trans. Med. Imaging* **20**(12), 1215–1227.

Muramatsu, C., Hara, T., Endo, T. & Others (2016), 'Breast mass classification on mammograms using radial local ternary patterns', *Computers in biology and medicine* **72**, 43–53.

Muramatsu, C., Zhang, M., Hara, T. & Others (2014), Differentiation of malignant and benign masses on mammograms using radial local ternary pattern, *in* H. Fujita, T. Hara & C. Muramatsu, eds, 'Breast Imaging', Vol. 8539 of *Lecture Notes in Computer Science*, Springer, pp. 628–634.

Nakagawa, Y. & Rosenfeld, A. (1979), 'Some experiments on variable thresholding', *Pattern Recognition* **11**(3), 191 – 204.

Nanni, L. & Lumini, A. (2008), 'A reliable method for cell phenotype image classification', *Artif. Intell. Med.* **43**(2), 87–97.

NAS (2008), 'BreastScreen Australia national accreditation standards: BreastScreen quality improvement program. Canberra, Australia', *Australian Government Department of Health and Ageing* p. 234.

Nascimento, M. Z., Martins, A. S., Neves, L. A. et al. (2013), 'Classification of masses in mammographic image using wavelet domain features and polynomial classifier', *Expert Systems with Applications* **40**(15), 6213 – 6221.

NBCF (2016), 'National Breast Cancer Foundation', `http://www.nationalbreastcancer.org/diagnostic-mammogram`. (Accessed: 17 April 2018).

NHS (2012), 'NHS Breast Screening Programme 2012 Annual Review', `http://www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp-annualreview2012.pdf`. (Accessed: 20 July 2017).

NHS (2018), 'Breast cancer screening UK', `https://www.nhs.uk/conditions/breast-cancer-screening/`. (Accessed: 17 April 2018).

Niu, Y., Wu, X. & Shi, G. (2016), 'Image enhancement by entropy maximization and quantization resolution upconversion', *IEEE Transactions on Image Processing* **25**(10), 4815–4828.

Nock, R. & Nielsen, F. (2004), 'Statistical region merging', *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(11), 1452–1458.

Obenauer, S., Sohns, C., Werner, C. et al. (2006), 'Impact of breast density on computer-aided detection in full-field digital mammography', *Journal of Digital Imaging* **19**(3), 258–263.

Ojala, T., Pietikainen, M. & Harwood, D. (1996), 'A comparative study of texture measures with classification based on feature distributions', *Pattern Recognition* **29**(1), 51–59.

Ojala, T., Pietikainen, M. & Maenpaa, T. (2002), 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987.

Oliver, A., Freixenet, J., Marti, J. et al. (2010), 'A review of automatic mass detection and segmentation in mammographic images', *Medical Image Analysis* **14**(2), 87–110.

Otsu, N. (1979), 'A threshold selection method from gray-level histograms', *IEEE Transactions on Systems, Man, and Cybernetics* **9**(1), 62–66.

Oztan, B., Shubert, K. R., Bjornsson, C. S. et al. (2013), Biologically-driven cell-graphs for breast tissue grading, *in* '2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)', IEEE, pp. 137–140.

Pandey, A., Yadav, A. & Bhateja, V. (2012), 'Contrast Improvement of Mammographic Masses Using Adaptive Volterra Filter', *Proceedings of the Fourth International Conference on Signal and Image Processing (ICSIP)* pp. 583–593.

Panetta, K. A., Wharton, E. J. & Agaian, S. S. (2008), 'Human visual system-based image enhancement and logarithmic contrast measure', *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **38**(1), 174–188.

Panetta, K., Zhou, Y., Agaian, S. et al. (2011), 'Nonlinear unsharp masking for mammogram enhancement', *IEEE Transactions on Information Technology in Biomedicine* **15**(6), 918–928.

Parker-Pope, T. (2008), 'Mammograms, New and Old', `https://well.blogs.nytimes.com/2008/04/10/mammograms-new-and-old/`. (Accessed: 17 April 2018).

Pawar, M. M., Talbar, S. N. & Dudhane, A. (2018), 'Local Binary Patterns Descriptor Based on Sparse Curvelet Coefficients for False-Positive Reduction in Mammograms', *Journal of healthcare engineering* **2018**.

Petrick, N., Chan, H.-P., Sahiner, B. et al. (1999), 'Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms', *Medical Physics* **26**(8), 1642–1654.

Pietikainen, M., Hadid, A., Zhao, G. et al. (2011), *Computer Vision Using Local Binary Patterns*, Springer Verlag, London.

Pietikainen, M. & Zhao, G. (2015), Two decades of local binary patterns: A survey, *in* 'Advances in Independent Component Analysis and Learning Machines', Academic Press, pp. 175–210.

Pietikäinen, M., Ojala, T. & Xu, Z. (2000), 'Rotation-invariant texture classification using feature distributions', *Pattern Recognition* **33**(1), 43–52.

Pisano, E. D., Cole, E. B., Hemminger, B. et al. (2000), 'Image Processing Algorithms for Digital Mammography: A Pictorial Essay', *RadioGraphics* **20**(5), 1479—-1491.

Pisano, E. D., Gatsonis, C. & et al, E. H. (2005), 'Diagnostic performance of digital versus film mammography for breast-cancer screening', *N Engl J Med* **353**, 1773–1783.

Pisano, E. D., Hendrick, R. E., Yaffe, M. et al. (2007), 'Should breast imaging practices convert to digital mammography? a response from members of the DMIST executive committee', *Radiology* **245**(1), 12–13.

Pizer, S. M., Amburn, E. P., Austin, J. D. et al. (1987), 'Adaptive histogram equalization and its variations', *Comput. Vision Graph. Image Process* **39**(3), 355–368.

Posso, M., Puig, T., Carles, M. et al. (2017), 'Effectiveness and cost-effectiveness of double reading in digital mammography screening: a systematic review and meta-analysis', *European Journal of Radiology* **96**, 40–49.

Pow, R. E., Mello-Thoms, C. & Brennan, P. (2016), 'Evaluation of the effect of double reporting on test accuracy in screening and diagnostic imaging studies: a review of the evidence', *Journal of Medical Imaging and Radiation Oncology* **60**(3), 306–314.

Pun, T. (1980), 'A new method for grey-level picture thresholding using the entropy of the histogram', *Signal Processing* **2**(3), 223 – 237.

Quinlan, J. R. (1986), 'Induction of decision trees', *Machine learning* **1**(1), 81–106.

Rabidas, R., Midya, A. & Chakraborty, J. (2018*a*), 'Neighborhood structural similarity mapping for the classification of masses in mammograms', *IEEE journal of biomedical and health informatics* **22**(3), 826–834.

Rabidas, R., Midya, A. & Chakraborty, J. (2018*b*), 'Neighborhood structural similarity mapping for the classification of masses in mammograms', *IEEE Journal of Biomedical and Health Informatics* **22**(3), 826–834.

Raghavendra, U., Acharya, U. R., Fujita, H. & Others (2016), 'Application of gabor wavelet and locality sensitive discriminant analysis for automated identification of breast cancer using digitized mammogram images', *Applied Soft Computing* **46**, 151–161.

Rahmati, P., Hamarneh, G., Nussbaum, D. et al. (2010), A New Preprocessing Filter for Digital Mammograms, *in* 'Image and Signal Processing', Vol. 6134 of *Lecture Notes in Computer Science*, Springer, pp. 585–592.

Rangayyan, R. M. (2005), *Biomedical Image Analysis*, CRC Press, Boca Raton, FL.

Rangayyan, R. M., Ayres, F. J. & Desautels, J. L. (2007), 'A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs', *Journal of the Franklin Institute* **344**, 312–348.

Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. E. L. et al. (1997), 'Measures of acutance and shape for classification of breast tumors.', *IEEE Transactions on Medical Imaging* **16**(6), 799–810.

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

Ravela, S. & Manmatha, R. (1997), Retrieving images by similarity of visual appearance, *in* '1997 Proceedings IEEE Workshop on Content-Based Access of Image and Video Libraries', pp. 67–74.

Rote, G. (1991), 'Computing the minimum hausdorff distance between two point sets on a line under translation', *Information Processing Letters* **38**(3), 123–127.

Sahiner, B., Chan, H. P., Petrick, N. et al. (1998), 'Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis', *Med. Phys.* **25**(4), 516–26.

Sahiner, B., Chan, H. P., Petrick, N. et al. (2001), 'Improvement of mammographic mass characterization using spiculation measures and morphological features', *Med. Phys* **28**(7), 1455–1465.

Sahoo, P., Soltani, S. & Wong, A. (1988), 'A survey of thresholding techniques', *Computer Vision, Graphics, and Image Processing* **41**(2), 233 – 260.

Sardanelli, F., Aase, H. S., Álvarez, M. et al. (2017), 'Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey', *European radiology* **27**(7), 2737–2743.

Sato, M., Kawai, M., Nishino, Y. et al. (2014), 'Cost-effectiveness analysis for breast cancer screening: double reading versus single+ cad reading', *Breast Cancer* **21**(5), 532–541.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012), 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods* **9**(7), 671–676.

Shannon, C. E. (1948), 'A mathematical theory of communication', *The Bell System Technical Journal* **27**(3), 379–423.

Shelda, M. & Ravishankar, M. (2013), 'Optimized Histogram Based Contrast Limited Enhancement for Mammogram Images', *ACEEE International Journal on Information Technology* **3**(1), 66–71.

Siegel, R., Ma, J., Zou, Z. et al. (2014), 'Cancer statistics, 2014', *A Cancer Journal for Clinicians* **64**(1), 9–29.

Singh, S. & Bovis, K. (2005), 'An Evaluation of Contrast Enhancement Techniques for Mammographic Breast Masses', *IEEE Transactions on Information Technology in Biomedicine* **9**(1), 109–119.

Sonka, M., Hlavac, V. & Boyle, R. (2007), *Image Processing, Analysis, and Machine Vision*, Thomson-Engineering.

Stark, J. A. (2000), 'Adaptive image contrast enhancement using generalizations of histogram equalization', *IEEE Transactions on Image Processing* **9**(5), 889–896.

Stephen, K., James, D., McCluggage, G. et al. (2000), 'An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (cin)', *The Journal of Pathology* **192**(3), 351–362.

Suckling, J., Parker, J., Dance, D. et al. (1994), 'The Mammographic Image Analysis Society digital mammogram database', *In Exerpta Medica, International Congress Series* **1069**, 375–378.

Suhail, Z., Hamidinekoo, A., Denton, E. R. et al. (2017), A texton-based approach for the classification of benign and malignant masses in mammograms, *in* 'Annual Conference on Medical Image Understanding and Analysis', Springer, pp. 355–364.

Sundaram, M., Ramar, K., Arumugam, N. et al. (2011), 'Histogram modified local contrast enhancement for mammogram images', *Applied Soft Computing* **11**(8), 5809–5816.

Susukida, H., Ma, F. & Bajger, M. (2008), 'Automatic tuning of a graph-based image segmentation method for digital mammography applications', *5th IEEE International Symposium on Biomedical Imaging* pp. 89–92.

Tabar, L. & Dean, P. B. (1982), 'Mammographic parenchymal patterns risk indicator for breast cancer', *JAMA* **247**(2), 185–189.

Tai, S. C., Chen, Z. S. & Tsai, W. T. (2014), 'An automatic mass detection system in mammograms based on complex texture features', *IEEE Journal of Biomedical and Health Informatics* **18**(2), 618–627.

Tan, X. & Triggs, B. (2010), 'Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions', *IEEE Transactions on Image Processing* **19**(6), 1635–1650.

Tang, J., Rangayyan, R. M., Xu, J. et al. (2009), 'Computer-aided detection and diagnosis of breast cancer with mammography : recent advances', *IEEE Transactions on Information Technology in Biomedicine* **13**(2), 236–251.

Theodoridis, S. & Koutroumbas, K. (2003), *Pattern Recognition*, 3rd ed. edn, Elsevier.

Tice, J. A., O'Meara, E. S., Weaver, D. L. et al. (2013), 'Benign breast disease, mammographic breast density, and the risk of breast cancer', *J Natl Cancer Inst* **105**(14), 1043–1052.

Torrents-Barrena, J., Puig, D., Ferre, M. et al. (2014), Breast masses identification through pixel-based texture classification, *in* H. Fujita, T. Hara & C. Muramatsu, eds, 'Breast Imaging', Vol. 8539 of *Lecture Notes in Computer Science*, Springer, pp. 581–588.

Tuceryan, M. & Jain, A. K. (1993), *Texture Analysis*, World Scientific Publishing Co.

Unay, D. & Ekin, A. (2008), Intensity versus texture for medical image search and retrival, *in* '2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro', pp. 241–244.

Ursin, G., Ma, H., AH, W. et al. (2003), 'Mammographic density and breast cancer in three ethnic groups', *Cancer Epidemiol Biomarkers* **12**(4), 332–340.

Ushizima, D. M., Bianchi, A. G. C. & Carneiro, C. M. (2014), 'Segmentation of subcellular compartments combining superpixel representation with voronoi diagrams', *Overlapping Cervical Cytology Image Segmentation Challenge-IEEE ISBI* pp. 1–2.

Vacek, P. M. & Geller, B. M. (2004), 'A prospective study of breast cancer risk using routine mammographic breast density measurements', *Cancer Epidemiol Biomarkers* **13**(5), 715–22.

Vallez, N., Bueno, G., Deniz, O. et al. (2014), 'Breast density classification to reduce false positives in CADe systems', *Computer Methods and Programs in Medicine* **113**, 569–584.

Vapnik, V. & Lerner, A. (1963), 'Pattern recognition using generalized portrait method', *Automation and Remote Control* **24**(3), 774–780.

Varela, C., Tahoces, P. G., Mendez, A. J. et al. (2007), 'Computerized detection of breast masses in digitized mammograms', *Computers in Biology and Medicine* **37**(2), 214–26.

Varma, M. & Zisserman, A. (2003), Texture classification: are filter banks necessary?, *in* '2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings', Vol. 2, pp. II–691–8 vol.2.

Wan, M., Gu, G., Qian, W. et al. (2018), 'Particle swarm optimization-based local entropy weighted histogram equalization for infrared image enhancement', *Infrared Physics and Technology* **91**, 164 – 181.

WCRFI (2017), 'World Cancer Research Fund International, breast cancer statistics', `https://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/breast-cancer-statistics`. (Accessed: 20 July 2017).

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

Weedon-Fekjær, H., Tretli, S. & Aalen, O. O. (2010), 'Estimating screening test sensitivity and tumour progression using tumour size and time since previous screening', *Statistical methods in medical research* **19**(5), 507–527.

Wei, C.-H., Li, Y. & Li, C.-T. (2007), Effective extraction of gabor features for adaptive mammogram retrieval, *in* '2007 IEEE International Conference on Multimedia and Expo', pp. 1503–1506.

Wei, D., Chan, H.-P., Petrick, N. et al. (1997), 'False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis', *Medical Physics* **24**(6), 903–914.

Wei, J., Chan, H.-P., Lu, Y. et al. (2012), Breast parenchymal pattern (bpp) analysis: comparison of digital mammograms and breast tomosynthesis, *in* 'International Workshop on Digital Mammography', Springer, pp. 514–520.

Weszka, J. S. (1978), 'A survey of threshold selection techniques', *Computer Graphics and Image Processing* **7**(2), 259 – 265.

Wolfe, J. N. (1976*a*), 'Breast patterns as an index of risk for developing breast cancer', *AJR Am J Roentgenol* **126**(6), 1130–1137.

Wolfe, J. N. (1976*b*), 'Risk for breast cancer development determined by mammographic parenchymal pattern', *Cancer* **37**(5), 2486–2492.

Wu, S., Yu, S., Yang, Y. et al. (2013), 'Feature and Contrast Enhancement of Mammographic Image Based on Multiscale Analysis and Morphology', *Computational and Mathematical Methods in Medicine* **2013**, 1–8.

Xi-Zhao, L. (2014), Breast cancer risk assessment using mammographic image texture analysis, PhD thesis, Flinders University, Australia.

Yaghjyan, L., Colditz, G. A. & Rosner, B. (2013), 'Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to the time since the mammogram', *Cancer Epidemiol Biomarkers* **22**(6), 1110–7.

Yang, Q., Gong, D. & Tang, X. (2005), *Analysis and Modelling of Faces and Gestures: Second International Workshop, AMFG 2005, Beijing, China, October 16, 2005. Proceedings*, Springer Berlin Heidelberg, Berlin, Heidelberg, chapter Modeling Micro-patterns for Feature Extraction, pp. 2–16.

Yener, B. (2016), 'Cell-graphs: Image-driven modeling of structure-function relationship', *Commun. ACM* **60**(1), 74–84.

Zhang, H., Fritts, J. E. & Goldman, S. A. (2004), An entropy-based objective evaluation method for image segmentation, *in* 'Proc. of SPIE: Storage and Retrieval Methods and Applications for Multimedia', Vol. 5307, pp. 38–49.

Zhang, J. & Tan, T. (2002), 'Brief review of invariant texture analysis methods.', *Pattern Recognition* **35**(3), 735–747.

Zhang, L., Chu, R., Xiang et al. (2007), *Face Detection Based on Multi-Block LBP Representation*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 11–18.

Zhang, W., Shan, S., Gao, W. et al. (2005), Local Gabor Binary Pattern Histogram Sequence (LGBPHS): a novel non-statistical model for face representation and recognition, *in* 'Tenth IEEE International Conference on Computer Vision, 2005', Vol. 1, pp. 786–791.

Zhen, L. & Chan, A. K. (2001), 'An artificial intelligent algorithm for tumor detection in screening mammogram', *IEEE transactions on medical imaging* **20**(7), 559–567.

Zheng, Y. (2010), 'Breast cancer detection with gabor features from digital mammograms', *algorithms* **3**(1), 44–62.

Zucker, S. W. (1976), 'Toward a model of texture', *Computer Graphics and Image Processing* **5**(2), 190 – 202.