Detection & Modelling of the Distribution of Linear Structures in Mammographic Images

Edward Michael Hadley

Department of Computer Science
Aberystwyth University

April
2013

This thesis is submitted in candidature for the degree of
Doctor of Philosophy.

Declaration & Statements

Declaration

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date .....

Statement 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote.
Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date .....

Statement 2

I hereby give consent for my thesis, if accepted, to be made available for photocopying and for inter–library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....

Abstract

Mammographic risk assessment is concerned with estimating the probability of a woman developing breast cancer. The aim is to improve the likelihood of early detection of breast cancers. The leading factor in determining risk is breast density, which has been shown to be the most accurate measure of mammographic risk, however more recently it has been suggested that the density (and possibly the distribution) of linear structures such as ducts and blood vessels within the breast are also related to mammographic risk. The purpose of this project is to investigate those relationships and the possibility of including this information in an automated risk assessment system.

A methodology is developed for detecting the linear structures in 2D mammograms. This information is used to calculate the density of linear structures which is used in a risk classifier. Results show that a classifier based on the density of linear structures outperforms a classifier based on breast density (64% correct BIRADS classification using linear density compared with 53% using breast density), and that a classifier combining both factors outperforms both individual classifiers (74% correct BIRADS classification), suggesting that linear density is related to risk and provides useful information for risk assessment.

The investigation in to the distribution of linear structures focusses on 3D tomosynthesis images. The linear structure detection methodology is developed for use in 3D, and a graph representation of the linear structures is extracted. Information from this graph relating to the distribution of linear structures is used for classification. The results of this classification (79% correct BIRADS classification) suggest that the distribution of linear structures is also related to risk and that this information provides additional risk–related information useful for risk assessment.

Acknowledgements

I would like to thank all of the people who helped and supported me during this research project, which continued for far longer than anyone imagined when it began.

I would like to thank my supervisor, Professor Reyer Zwiggelaar, for his academic excellence, support and guidance throughout this project. I would also like to thank the members of the Vision, Graphics and Visualisation Group in Aberystwyth for their support and feedback.

I am very grateful to my wife, Emma, for her support, understanding and encouragement, and for putting up with me spending endless evenings working on the computer rather than doing everything that I should have been doing. I am thankful for my children, Nathaniel and Sebastian, who had not been born when I began this project but have been a source of happiness and joy.

I would also like to thank my parents, Michael and Helen, my brother, Philip, and sister, Rachel, for their encouragement and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Each year, an estimated 1.38 million new cases of breast cancer are diagnosed worldwide, leading to more than 450,000 deaths [1, 2]. It is amongst the most common forms of cancer and is the most common form in women [3, 4, 5]. In the United States, breast cancer forms 26% of cancers diagnosed in women and is the cause of 15% of cancer–related deaths [3]. In Europe, breast cancer has been estimated to be the most common form of cancer with approximately 430,000 new cases annually, representing 13.5% of all cancer cases [6].

Regular screening for breast cancer has become common in many developed countries and has been shown to significantly reduce mortality rates by detecting more cancers early when they are at their most treatable [7]. Mammographic risk assessment involves estimating a patient's risk of developing breast cancer and aims to further improve the early detection rate of breast cancer by allowing for improved targeting of resources towards those at greatest risk [8]. Those considered to be at higher risk can be offered greater surveillance, lifestyle recommendations

or preventative measures.

Mammographic risk assessment is often performed manually using a model such as the Gail model [9, 10] which uses personal characteristics to produce a risk estimate. Such methods are time–consuming and error–prone due to their dependence on the accurate reporting of risk factors by the patient. As such, strong interest exists in the possibility of developing automated risk assessment methodologies and has been assisted by the development of full–field digital mammography as a replacement for traditional screen–film mammography which introduces computers to the capture–display sequence allowing for analysis algorithms [11, 12].

Breast density has been shown to be a strong predictor of mammographic risk and is the most common quantitative method of measuring risk [13, 14, 15, 16, 17, 18, 19]. Most automated risk assessment methods have focussed on the area or volume of dense breast tissue as the principal factor. However, more recently, Tabár *et al.* have suggested that the density and possibly the distribution of linear structures (ducts, blood vessels, etc.) in the breast are also correlated with mammographic risk [20, 21, 22]. The incorporation of this information into an automated risk classifier could lead to more accurate risk assessment models which in turn have the potential to improve the early detection of breast cancer.

1.2 Research Objectives

There are six principal research objectives within the project.

1. **Select a suitable mammographic line detection methodology.**

   The relevant line detection literature is reviewed and a specific approach for detecting linear structures in 2D mammograms is implemented. As a pre-

cursor to classification, initial experimental results are obtained to assess whether it is possible to distinguish between the detected linear density in images of different risk categories.

Figure 1.1 shows some typical mammograms used during the project.

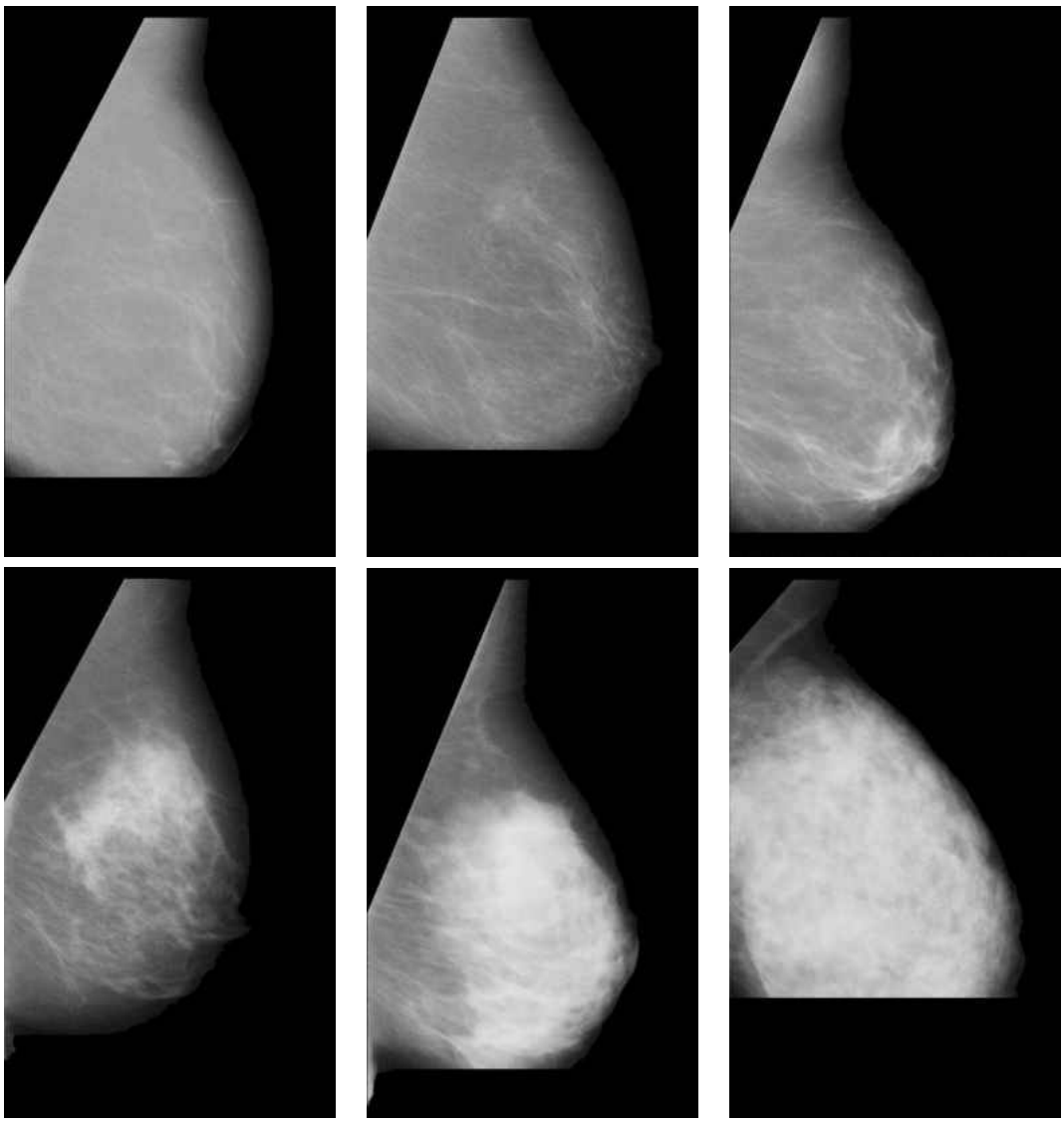


Figure 1.1: Typical 2D mammograms from the MIAS database [23]. Mammograms are ordered according to risk, where the top–left shows the lowest risk mammogram and the bottom–right shows the highest risk mammogram.

2. **Investigate the relationship between the density of linear structures in 2D mammograms and mammographic risk.**

   This investigation will assess whether the density of linear structures can be used to automatically identify mammographic risk, and how its performance

compares with a more conventional density–based approach. The density of linear structures detected from the images is used to classify images according to commonly used mammographic risk assessment metrics.

3. **Detect mammographic linear structures in raw tomosynthesis images.**

   Tomosynthesis images provide the potential for significant improvements over conventional mammography [24, 25, 26]. Raw tomosynthesis images consist of a sequence of low–dose x–ray images taken in an arc around the breast, however the x–ray dose used per image is significantly lower than that used for conventional 2D mammography. This results in poorer contrast and so initial experiments are conducted in order to assess the viability of using linear structure detection techniques in tomosynthesis images. The purpose of this objective is to investigate if the linear structures in raw tomosynthesis images can be detected automatically and can be correlated with conventional mammograms.

4. **Detect mammographic linear structures in reconstructed tomosynthesis images.**

   A methodology is be developed for the automatic detection of mammographic linear structures in reconstructed (three–dimensional) tomosynthesis images. This is based on the methodology used for detecting linear structures in 2D mammograms.

5. **Develop a model of the distribution of linear structures.**

   A three–dimensional network representation of the distribution of mammographic linear structures within the breast is developed, where edges of the network represent linear structures and nodes represent end or branching points. This can be derived from the results from objective 4.

6. **Investigate how the distribution of mammographic linear structures might be related to risk.**

Tabár *et al.* have proposed a mammographic risk assessment model based on four structural components, where the relative proportions of each component is linked to the risk of developing breast cancer [20, 21, 22]. One of the four structural components is linear density. The main purpose of this scientific aim is to investigate if automatic methods can be used to correlate the distribution of linear structures (SA5) with mammographic risk.

1.3 Contribution

The principal purpose of this project is to investigate whether the inclusion of linear structure information might improve an automated risk classifier. Automated risk classification has the potential to be incorporated into a digital mammographic screening programme, providing risk information relating to each patient. This information could be used to target limited resources at those patients most at risk of developing breast cancer with the aim of increasing early detection rates overall [8].

During the project, a methodology is developed for detecting the linear structures in 2D mammograms. This information is used to calculate the density of linear structures which is used in a risk classifier. Results show that a classifier based on the density of linear structures outperforms a classifier based on breast density (64% correct BIRADS classification using linear density compared with 53% using breast density), and that a classifier combining both factors outperforms both individual classifiers (74% correct BIRADS classification), suggesting that linear density is related to risk and provides useful information for risk assessment.

The investigation into the distribution of linear structures focusses on 3D tomosynthesis images, which have been suggested as a replacement for conventional

2D mammography in breast screening [24, 25, 26]. The linear structure detection methodology is developed for use in 3D, and a graph representation of the linear structures is extracted. Information from this graph relating to the distribution of linear structures is used for classification. The results of this classification (79% correct BIRADS classification) suggest that the distribution of linear structures is also related to risk and that this information provides additional risk–related information useful for risk assessment.

In summary, the results of this project suggest that linear structure information could be incorporated into an improved automated risk classifier which could provide useful risk–related information as part of a mammographic screening programme.

Chapter 2

Review of Literature

2.1 Mammographic Risk Assessment

Mammographic risk assessment is concerned with the probability of women developing breast cancer. The aim is to improve the likelihood of early detection of breast cancers, which is critical to maximise treatment success and long–term survival. The principal factor in determining risk is breast density, suggested by Wolfe as early as 1976 [14] and has been shown to be amongst the most accurate measures of mammographic risk [13]. Several image–based risk assessment metrics exist, including Wolfe [14], Breast Imaging Reporting and Data System (BIRADS) Density Classification [27], Boyd SCC [28] and Tabár [22].

The BIRADS and Boyd metrics are entirely density–based. Boyd defined a system of six class categories (SCC) of progressively increasing breast density, where class 1 corresponds to the least dense and lowest risk group and class 6 relates to the most dense and highest risk [28]. The BIRADS metric defines four categories of increasing breast density, where class I represents the least dense and lowest risk and class IV represents the most dense and highest risk [27].

The Wolfe and Tabár metrics are based on breast density and the parenchymal patterns present within the breast. Wolfe defined four categories of mammographic risk, based primarily on the presence of dense or fatty tissue and the ductal pattern. Wolfe's class 1 represents the lowest risk and class 4 represents the highest risk [14]. Tabár proposed a system whereby the breast is composed of four building blocks: nodular density, linear density, homogeneous fibrous tissue and radiolucent adipose tissue. The relative proportions of these four building blocks determine classification in a system of five Patterns, where Patterns I–III are related to low mammographic risk and Patterns IV–V correspond to high risk [22].

Comparison between Wolfe, Boyd, BI-RADS and Tabár based mammographic risk assessment by Muhimmah *et al.* [29] investigated these four risk assessment metrics and the problems of inter– and intra–observer variability. The study used a set of 321 mammograms which were classified independently by three expert radiologists according to each of the four risk metrics. The conclusion was that the Wolfe, Boyd and BIRADS classifications were strongly correlated (Spearman's correlation $(r_S) = 0.908 - 0.929$) whilst the Tabár classification was not well correlated with the other three metrics ($r_S = 0.408 - 0.459$). It was discovered that the difference was almost entirely due to Tabár Pattern I and that when this pattern was removed, the remaining patterns demonstrated a very strong correlation with the other three classes ($r_S = 0.93 - 0.96$).

The authors did not attempt to explain why Tabár Pattern I may not correlate well with the other risk metrics other than to say that further investigation is required in order to determine its relationship to risk.

Quantitative assessment of mammographic breast density: relationship with breast cancer risk by Harvey and Bovbjerg [30] argued that there is a moderate association between breast density and breast cancer risk. The paper reviewed several studies across a variety of settings and populations and concluded that

breast density is a moderate independent risk factor for breast cancer. It says that only genetic abnormalities, age, the presence of prior breast cancer or biopsy findings showing lobular carcinoma in situ have a similar or greater association with breast cancer risk. The authors stated that the use of breast density in risk assessment is hindered by its inability to be measured in a reproducible, standardised and automated fashion. Breast density, unlike other risk factors, can be influenced by effects such as hormone levels, alcohol intake and diet. The paper finishes by saying that further research is required to determine whether reducing breast density results in lower risk.

Another factor that has been investigated is the relationship between breast asymmetry and risk. *Breast asymmetry and a predisposition to breast cancer* by Scutt *et al.* [31] shows that breast asymmetry is a contributing factor to breast cancer. The described study used 252 asymptomatic women who had normal mammography, but went on to develop breast cancer, and 252 age–matched healthy controls who remained free of breast cancer for the duration of the study. The control group were chosen randomly from lists of suitable candidates in each age stratum. The mean and standard deviation of the cancer and control groups were very similar.

The study found that the women who went on to develop breast cancer had greater breast asymmetry than the control group, leading to the conclusion that breast asymmetry is a factor that can be considered in mammographic risk assessment. Since breast asymmetry affects mammographic risk, it may be possible for an automated system to compare the density and distribution of linear structures between the left and right breasts.

2.1.1 Density, Boyd & BI-RADS

Breast density has been established as the leading indicator of mammographic risk. Boyd *et al.* in *Mammographic Density and the Risk and Detection of Breast Cancer* [19] stated that breast density has a strong association with mammographic risk. The study involved 1112 images from breast cancer patients and 1112 paired control images where no breast cancer developed.

The images were placed into one of six Boyd density categories based on their percent density. Classes 1–6 correspond to 0%, $< 10\%$, $10- < 25\%$, $25- < 50\%$, $50- < 75\%$ and $\geq 75\%$ density respectively. The authors discovered that cases in Boyd SCC classes 3, 4, 5 and 6 are respectively 1.8, 2.1, 2.4 and 4.7 times more likely to develop breast cancer than those in classes 1 and 2.

Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis by McCormack & dos Santos Silva [32] provided a meta–analysis of 42 publications on mammographic patterns in relation to breast cancer risk. The authors analysed studies relating breast cancer risk to percent density, Wolfe, BIRADS and Tábar classifications. The conclusions concurred with the studies by Boyd *et al.* that breast density is strongly correlated with mammographic risk. The correlation with BIRADS and Wolfe classification (which are largely based on breast density) were also strong.

The study [32] showed the results of the meta–analysis of breast cancer risk related to Wolfe, BIRADS and Tábar classifications. The analysis according to Wolfe classifications has combined the four classes into low risk (classes 1 and 2) and high risk (classes 3 and 4) groups. The studies were divided into incidence studies in the general population (where assessment was made with a negative mammogram and cancer was diagnosed subsequently), prevalence studies in the general population (where assessment was made on the contralateral breast at

diagnosis) and studies in symptomatic populations.

In all three types of study, the combined results showed an increased risk of breast cancer in the 'high risk' group. This was strongest in the symptomatic population (6.9 times more likely than the low risk group), than in the studies in the general population (incidence and prevalence studies showed 1.86 and 1.44 times more likely than the low risk group, respectively). However, there was a high degree of variation between studies, particularly in studies of the general population.

Analysis by BIRADS class showed strong correlation with mammographic risk. The combined results for cases in BIRADS classes 2, 3 and 4 indicated a likelihood of developing breast cancer of 2.04, 2.81 and 4.08 times that of BIRADS class 1, respectively.

A notable result was the analysis related to Tábar patterns. Since Tábar patterns I–III should indicate a low risk of breast cancer and patterns IV–V should indicate a higher risk, it would be expected that when using pattern I as a reference, that patterns II–III should not show a substantially higher risk, whereas patterns IV–V should indicate a higher risk [22]. The results observed, however indicate a lower risk for pattern II, a high risk for patterns III–IV and an approximately equal risk for pattern V.

The study by Muhimmah *et al.* discussed above indicated a possible anomaly with Tábar pattern I, whereby patterns II–V correlate well with other metrics but pattern I does not [29]. It may have been more relevant to use another pattern, such as pattern II as the reference, however the results would still suggest that pattern III indicates relatively high risk and pattern V indicates a lower risk. It should also be noted that the analysis related to Tábar pattern was based on a single study, whereas the other analyses were all based on multiple studies [32].

The study by McCormack & dos Santos Silva [32] showed the results of the meta–analysis of breast cancer risk related to breast percent density. The combined results showed a clear correlation between breast percent density and mammographic risk. The results for incidence cases with densities of $5-24\%$, $25-49\%$, $50-74\%$ and $75\%+$ were respectively 1.79, 2.11, 2.92 and 4.64, times more likely to develop breast cancer than cases of less than 5% density. These figures closely resemble those discussed previously by Boyd *et al.* [19], supporting the reliability of these figures and the conclusion of a strong correlation between risk and density.

Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk by Harvey and Bovbjerg provided a review of studies investigating the relationship between breast density and risk [30]. The review investigated 12 studies conducted since 1982 using quantitative methods to investigate the relationship between density and risk and all 12 reported a moderate to strong positive association between increased cancer risk and higher levels of breast density. The authors discussed whether the association between density and risk is due to other risk factors for which density is a marker or whether density is an independent risk factor. Unlike other risk factors such as family history, breast density can be influenced by hormones or by lifestyle factors such as diet. While weight, body mass index, age, menopausal status, age at first birth, nulliparity, family history, hormone use, and previous breast biopsy may all influence breast density, the authors stated that breast density is consistently correlated with risk even after adjustment for other variables associated with both density and breast cancer risk [33, 34], suggesting that density is an independent risk factor. Boyd *et al.* reviewed 9 of the 12 studies reviewed by Harvey and Bovbjerg and discovered that cancer risk increased proportionally with density in 8 of the 9 studies reviewed [35]. The authors discussed the theory of dense tissue masking, proposed by Egan and Mosteller [36], which suggests that the increased risk may be caused by dense tissue masking the early stages of abnormalities and thus delaying diagnosis. However, if this was the sole reason for the correlation between density

and risk, at some point in the future it would be expected that the incidence of breast cancer in women with dense and fatty breasts should become similar. The authors state that this theory alone cannot explain the correlation. Other studies have suggested that the masking effect leading to decreased mammographic sensitivity in dense breasts may contribute to mammographic failure, particularly in younger women [37].

Effect of Baseline Breast Density on Breast Cancer Incidence, Stage, Mortality, and Screening Parameters: 25-Year Follow-up of a Swedish Mammographic Screening by Chui *et al.* [38] described a large–scale study carried out over 25 years in Kopparberg County, Sweden, the first results of which were published in 1985 by Tabár *et al.* [39]. A total of 15,658 women were followed from 1977 to 2004 to ascertain breast cancer incidence and mortality. During the trial period, women in the active study population provided information on breast density while they underwent mammographic examination.

Breast density was classified as dense (Tabár patterns IV–V) or non–dense (Tabár patterns I–III). In addition, information on age at menarche, body mass index (BMI), age at first full-term pregnancy, and menopausal status was obtained through a questionnaire prior to screening.

The study [38] showed the results of the analysis of risk factors for breast cancer incidence. Whilst the results show that incidence is affected by many of the factors measured, the greatest effects are seen in the analysis of Tabár pattern/density where patterns IV–V are shown in the multivariate analysis as having an incidence rate 1.57 times that of patterns I–III. These results lend further support to density being a leading indicator of mammographic risk.

In summary, numerous studies have concluded that breast density is highly correlated with mammographic risk and can be used as an indicator of risk. The Boyd SCC classification and BIRADS density classification metrics are entirely

based on breast density and are intended to provide an indication of risk. Other risk assessment metrics, such as Wolfe and Tabár, incorporate breast density as well as other information.

2.1.2 Parenchymal Patterns, Wolfe & Tabár

It has been suggested that parenchymal patterns within the breast may also be related to breast cancer risk. *Wolfe Mammographic Parenchymal Patterns and Breast Cancer Risk* by Sickles [40] discussed Wolfe's classification metric [14], in particular the contribution of parenchymal pattern information.

Wolfe devised a system of four parenchymal pattern categories: N1, primarily fatty; P1, $\leq 25\%$ prominent ducts; P2, $> 25\%$ prominent ducts; and DY, dense fibroglandular tissue. Throughout this work these categories are generally referred to as 1, 2, 3 and 4 respectively. Wolfe suggested that class 4 represented a 37 times greater risk than class 1 [14]. Sickles argues that a combination of a young, primarily symptomatic study population, the choice of a 6-month exclusion interval, and the limited mammographic capability of the late 1960s and early 1970s make it likely that a very large proportion of the future cancers identified by Wolfe were in fact masked cancers already present, and that this would disproportionately affect dense–breasted women [40]. The result was that the future breast cancer risk stated by Wolfe was greatly overestimated, leading to controversy as to the extent that parenchymal patterns contribute to risk.

More recently, Tabár *et al.* have suggested that the parenchymal patterns within a breast may be linked to breast cancer risk [20, 22]. Tabár has devised a classification system based on four mammographic building blocks or tissue types. The relative proportions of these building blocks determine the pattern. One of Tabár's building blocks is linear density [22], suggesting that the density of linear structures within the breast is linked to risk.

2.1.3 Hereditary & Other Non–density Factors, Gail & Claus

Mammographic risk is known to be related to hereditary factors [41], and the Gail model includes hereditary factors in an estimate of mammographic risk [9] in addition to other risk factors such as the woman's age at first live birth, her age at menarche, the number of previous benign breast biopsies and whether or not atypical hyperplasia had been identified on any of the biopsies.

Gail Model for Prediction of Absolute Risk of Invasive Breast Cancer: Independent Evaluation in the FlorenceEuropean Prospective Investigation Into Cancer and Nutrition Cohort by Decarli *et al.* [42] investigated the sensitivity of the Gail model using large data sets from the US and Italy. The authors concluded that the Gail model is a good predictor of breast cancer, with observed incidence rates closely matching the expected incidence rates for both data sets.

Prospective Breast Cancer Risk Prediction Model for Women Undergoing Screening Mammography by Barlow *et al.* [43] investigated the addition of breast density with more traditional risk factors such as those used in the Gail model. The study used more than two million screening mammograms from one million women who had not previously been diagnosed with breast cancer. Within one year of the screening mammogram, 11,368 women had been diagnosed with breast cancer.

The authors used logistical regression analysis to estimate the probability of developing breast cancer within one year based on each screening mammogram. Since risk factors varied depending on menopausal status, the dataset was divided into premenopausal and postmenopausal women which were subjected to separate statistical analysis. The sets were then analysed according to a variety of individual risk factors in order to find a model that best predicted the likelihood of developing breast cancer. In each case, a random 75% of the dataset was used as

the training set and the remaining 25% was used as the validation set.

Results of particular interest have been reproduced in Tabs. 2.1 and 2.2 for premenopausal and postmenopausal women respectively. The results provide an interesting overview of a wide variety of mammographic risk factors and an indication of the significance of each factor.

Results shown in Tab. 2.1 indicate that in premenopausal women, the study found that breast density was the most significant risk factor, with BIRADS class 4 cases being almost 4 times as likely to develop breast cancer as BIRADS class 1 cases. Age and family history of breast cancer were also major risk factors, where women with two or more relatives diagnosed with breast cancer are more than twice as likely to develop breast cancer as those with no family history.

Table 2.2 shows results for postmenopausal women, where the most significant risk factor is age, followed closely by breast density. Family history was less significant in postmenopausal women than in premenopausal women. In both datasets, the most significant risk factor other than age was found to be breast density, supporting the work of Boyd *et al.* [19]. No factors other than age, breast density and (in premenopausal women) family history demonstrated a odds ratio of greater than 2.0 or less than 0.5. Family history of breast cancer was less significant than density in both cases, suggesting that density–based risk estimation has the potential to be more accurate than an estimation based on other factors.

Table 2.2: Table reproduced from Barlow *et al.* [43] showing multivariable model for predicting breast cancer within 1 year in postmenopausal women who have undergone previous screening.

Risk factor	**Odds ratio (95% CI)**
Age	
45–49	1.00 (referent)
50–54	1.33 (1.18 to 1.50)
55–59	1.96 (1.75 to 2.20)
60–64	2.27 (2.02 to 2.55)

Continued on next page

Continued from previous page

Risk factor	**Odds ratio (95% CI)**
65–69	2.47 (2.20 to 2.77)
70–74	2.79 (2.48 to 3.14)
75–79	3.03 (2.69 to 3.42)
80–84	3.33 (2.91 to 3.80)
Hispanic	
Non-Hispanic	1.00 (referent)
Hispanic	0.74 (0.66 to 0.81)
Unknown	0.94 (0.87 to 1.01)
Race	
White	1.00 (referent)
Asian/Pacific Islander	0.80 (0.71 to 0.90)
Black	1.10 (1.00 to 1.21)
Native-American/Alaskan Native	0.54 (0.42 to 0.70)
Other	0.97 (0.77 to 1.24)
Unknown	1.00 (0.92 to 1.09)
Body mass index, kg/m2	
< 25 (referent)	1.00 (referent)
25–29.99	1.14 (1.07 to 1.23)
30–34.99	1.28 (1.17 to1.40)
35 or greater	1.47 (1.30 to 1.65)
Missing	1.03 (0.97 to 1.10)
Age at birth of first child, y	
< 30	1.00 (referent)
≥ 30	1.21 (1.09 to 1.33)
Nulliparous	1.18 (1.09 to 1.27)
Unknown	1.02 (0.96 to 1.08)
Prior breast procedure	
No	1.00 (referent)
Yes	1.30 (1.24 to 1.36)
Unknown	1.06 (0.98 to 1.15)
First-degree family history of breast cancer	
No	1.00 (referent)
Yes, one relative	1.31 (1.24 to 1.39)
Yes, two or more relatives	1.66 (1.39 to 1.99)
Unknown	0.95 (0.89 to 1.02)
Current hormone therapy use	
No	1.00 (referent)
Yes	1.19 (1.13 to 1.24)
Unknown	1.13 (1.05 to 1.22)
Surgical menopause	
No	1.00 (referent)
Yes	0.84 (0.80 to 0.89)
Unknown	0.94 (0.89 to 1.00)
Previous mammographic outcome	
Negative	1.00 (referent)

Continued on next page

Continued from previous page

Risk factor	**Odds ratio (95% CI)**
Positive	1.69 (1.47 to 1.94)
Unknown	1.20 (1.14 to 1.26)
Breast density (BI-RADS)	
1. Almost entirely fat	1.00 (referent)
2. Scattered fibroglandular densities	2.09 (1.86 to 2.35)
3. Heterogeneously dense	2.95 (2.61 to 3.32)
4. Extremely dense	3.15 (2.72 to 3.66)
Unknown or different system	2.84 (2.52 to 3.21)

Another study that investigated the relationship of breast density with non-density risk factors is *Mammographic Density Correlation with Gail Model Breast Cancer Risk Estimates and Component Risk Factors* by Palomares *et al.* [44]. This study compared mammographic density with risk as calculated using the Gail model using a set of 99 participants of a clinical trial. The authors found that although most increases in breast density could be attributed to factors included in the Gail model, seven percent of density increases could not be explained using the Gail model factors. This suggests that breast density should be included in a risk model since there is information present in breast density that is not captured by the Gail model.

A limitation of the study was the size of the dataset. A relatively small percentage difference in a dataset of 99 may simply be due to chance, so for a more reliable conclusion a repeat of the experiment using a larger dataset would be beneficial.

Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme by Amir *et al.* [45] assessed several non–density based risk assessment models. The investigation used four published models (Gail, Claus, Ford and Tyrer–Cuzick) and one manual model based on the Claus model but adjusted for other risk factors.

Table 2.1: Table reproduced from Barlow *et al.* [43] showing multivariable model for predicting breast cancer within 1 year in premenopausal women who have undergone previous screening.

Risk factor	**Odds ratio (95% CI)**
Age	
35–39	1.00 (referent)
40–44	1.23 (0.96 to 1.56)
45–49	1.89 (1.49 to 2.41)
50–54	2.36 (1.84 to 3.02)
Prior breast procedure	
No	1.00 (referent)
Yes	1.47 (1.31 to 1.65)
Unknown	0.87 (0.69 to 1.10)
First-degree family history of breast cancer	
No	1.00 (referent)
Yes, one relative	1.54 (1.36 to 1.74)
Yes, two or more relatives	2.11 (1.32 to 3.36)
Unknown	0.89 (0.75 to 1.07)
Breast density (BI-RADS)	
1 (almost entirely fat) (referent)	1.00 (referent)
2 (scattered fibroglandular densities)	2.00 (1.26 to 3.18)
3 (heterogeneously dense)	3.34 (2.12 to 5.27)
4 (extremely dense)	3.93 (2.46 to 6.28)
Unknown/different system	3.31 (2.09 to 5.24)

The Gail model [9, 10] includes limited information on family history and focusses on non–genetic risk factors, whereas the Claus model [46] is based primarily on family history of breast cancer. Ford *et al.* [47] described a model based on family history and personal characteristics aimed at identifying the presence of any mutation of the *BRCA* genes. A more recent model described by Tyrer *et al.* [48] attempts to address many of the shortcomings of the other models by incorporating a wider combination of factors than any single earlier model, including extensive family history and other factors.

The study by Amir *at al.* used each risk model to estimate the lifetime and 10–year risk of developing breast cancer for each of a set of 3170 women. The number of expected cases of breast cancer was calculated using each model and these results were compared with the observed number of cases (64) to provide

an indication of accuracy. The results showed that the Tyrer–Cuzick and manual models were the most accurate, predicting 69.6 and 78.2 cases respectively. Although both methods slightly overestimated the risk, the difference was not statistically significant. The Gail, Claus and Ford methods all significantly underestimated risk, predicting 44.3, 48.6 and 42.3 cases respectively. This suggests that whilst a combination of family history factors and personal factors provides a reasonably accurate risk estimation, individually these factors do not provide sufficient information on which to make an accurate risk assessment. A larger study population would provide a more accurate assessment, since the small number of cases in this study led to wide confidence intervals.

Mammographic Breast Density and the Gail Model for Breast Cancer Risk Prediction in a Screening Population by Tice *et al.* [49] used a screening cohort of $81,777$ women to compare the accuracy of the Gail model with the BIRADS density classification. Within 5.1 years, 955 women had been diagnosed with breast cancer. The authors found that there was no difference in the accuracy of the Gail and BIRADS models, suggesting that breast density provides an equivalent level of accuracy as a risk metric. Interestingly, the Gail model with density added provided only a marginal improvement over each individual model. This may suggest that there is very little additional information present in the density estimation, since density can be affected by many of the risk factors present in the Gail model.

Viewed in conjunction with the study by Amir *at al.* [45], this may indicate that a combination of hereditary and other risk factors could provide a better indication of risk than density alone, although more research would need to be carried out in order to confirm this.

Some studies have investigated whether the risk of breast cancer can be effectively reduced by making changes affecting known risk factors. *Prevention of Breast Cancer in Postmenopausal Women: Approaches to Estimating and Reduc-*

ing Risk by Cummings *et al.* [50] provided a meta–analysis of studies investigating the relationship between various factors and breast cancer risk, and investigated the effects of exercise, weight management, healthy diet, moderate alcohol consumption, and fruit and vegetable intake on breast cancer risk.

The study concluded that a combination of risk factors including breast density and other factors provides the most accurate risk estimates, and the authors suggested that exercise, weight management, and reducing alcohol intake should be recommended to postmenopausal women to reduce breast cancer risk. However, increased intake of fruit and vegetables was not associated with a reduced risk of breast cancer.

This study has some limitations. Firstly, most of the data concerned only postmenopausal women, and provided little evidence as to whether such lifestyle changes in premenopausal women would reduce breast cancer risk. Further, many of the studies on which the analysis is based used self–evaluation to assess lifestyle factors, which introduced a potential source of error. Finally, whilst the studies demonstrated a link between these lifestyle factors and risk, they did not specifically assess the effect of changes in these factors, which would require a wider clinical study.

Breast Cancer Risk Reduction by Bevers *et al.* [41] investigated the effects of some risk reduction strategies but primarily focussed on more severe methods such as risk reduction surgery and pharmaceutical agents such as tamoxifen and raloxifene, however the authors also studied the effects of lifestyle factors on the reduction of breast cancer risk.

The authors concluded that whilst various studies have shown that moderate alcohol intake is associated with a $30\% - 50\%$ increased risk of breast cancer [51], the effect of a reduction in alcohol intake has not been well–studied.

Several studies were analysed that investigate the effects of exercise on breast cancer risk, all of which supported the association between increased exercise and reduced levels of risk. A study of 90,509 women found a relative risk of 0.62 (95% CI, 0.49–0.78) in women who performed at least 5 hours of vigorous exercise per week, compared with those who did not perform regular exercise [52].

A case–control study of 4538 case patients with newly–diagnosed breast cancer found that women with annual lifetime exercise activity levels exceeding the median activity level for active control subjects had a 20% lower breast cancer risk than inactive women [53]. Finally, a study of 45,631 women found that those who reported walking/hiking for at least 10 hours per week had a relative risk of 0.57 (95% CI, 0.34–0.95) compared with inactive women. As discussed, the evidence generally suggests an inverse association between exercise and risk, however it is not clear what effect an increase in the level of exercise would have [41].

Bevers *et al.* concluded that the association between diet and breast cancer risk is unclear. A trial of 48,835 women investigating the effects of a low fat diet on breast cancer risk in postmenopausal women over 8 years found no statistically significant reduction in risk [54]. The authors note, however, that such trials are notoriously difficult to maintain and assure compliance with the dietary requirements of the trial. Additionally, it is possible that diet may have more of an effect in younger women [41].

The authors finally assessed the effects of weight/body mass index and conclude that a substantial amount of evidence suggests an association between higher body mass index and increased breast cancer risk [41]. A study of 87,143 women concluded that women who gain at least $25kg$ since 18 years of age are 1.45 (95% CI, 1.27–1.66) times more likely to develop breast cancer than those who maintained their weight [55].

2.2 3D Breast Tomosynthesis

A number of studies have been published which provide a background for the need for advances in visualisation for breast cancer. *The false-negative mammogram* by Huynh *et al.* [56], investigated many of the problems in screening for breast cancer that lead to false-negative results (i.e. a cancer going undetected). The paper identified four principal factors that cause false-negative results: the inherent limitations of screen-film mammography, inadequate radiographic technique, subtle or unusual lesion characteristics, and interpretation error.

The limitations of screen-film mammography include restricted latitude and contrast, which can make tumours and microcalcifications difficult to detect, particularly in patients with dense fibroglandular tissue. Radiographic technique includes factors affecting the taking of x-ray images, such as positioning, compression and image exposure.

The paper stated that radiographers are more likely to miss lesions that are small, have a probably benign appearance, or are in an area where visualisation is difficult, and that common causes of interpretation error include factors such as suboptimal viewing conditions, outside distractions, lack of systematic approach, oversight of a subtle lesion due to the finding of a more obvious one, lack of knowledge, and inaccurate correlations of results with other studies.

Huynh *et al.* [56] stated that there is very little that can be done to overcome many of these limitations, however there have since been developments towards digital imaging and 3D imaging which could potentially help improve aspects such as contrast, and allow the radiologist to manipulate images on a computer system, altering quantities such as the contrast to show features more clearly. Three–dimensional imaging has the potential to overcome many of the limitations inherent in 2D imaging by removing problems such as overlapping dense tissue

masking an abnormality and overlapping dense tissue areas giving the false appearance of an abnormality [24, 57].

Digital breast tomosynthesis is a technique used to produce a three–dimensional image volume using x–ray images. The image is captured as a series of two–dimensional digital x–ray images taken at regular intervals in an arc around the patient (see Fig. 2.1). These images are then reconstructed using a algorithm such as filtered back–projection to form a three–dimensional volume [58, 59].

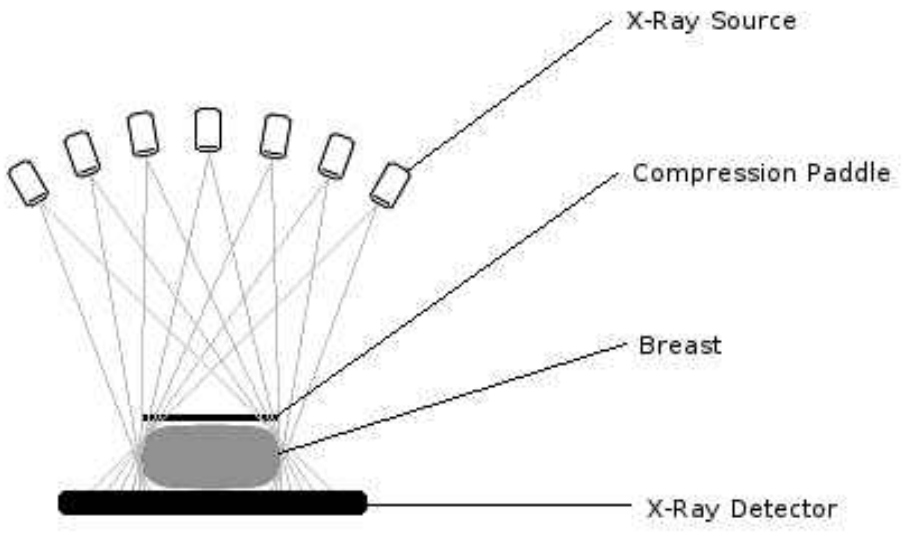


Figure 2.1: Diagram showing typical tomosynthesis acquisition. The x–ray source moves in an arc around the stationary breast and detector, capturing a sequence of x–ray images at regular intervals.

A principal advantage of tomosynthesis over other 3D imaging techniques such as magnetic resonance imaging (MRI) is cost. MRI images are expensive, require highly specialised equipment, and are time consuming to capture, making them unsuitable for general screening purposes. Tomosynthesis images can be captured using equipment similar to that used to capture 2D digital mammograms, and most breast tomosynthesis machines in current use are capable of both [59, 57].

The individual x–ray images captured during tomosynthesis acquisition are typically lower–dose than a 2D mammogram, so as to keep the overall exposure similar to a regular 2–image 2D screening session [57].

The three–dimensional nature of breast tomosynthesis provides the possibility to overcome many of the shortcomings of 2D mammography. Using 2D mammograms, problems occur when dense tissue in front or behind an abnormality creates a masking effect, frequently leading the reader to overlook the abnormality [57]. Similarly, overlapping regions of dense tissue can sometimes create the false effect of an abnormality, leading to unnecessary, stressful and often–painful follow–up procedures. Park *et al.* conclude that tomosynthesis has the potential to reduce the false–positive rate by providing the reader with better visibility of any potential areas of concern [57].

Some recent studies have conducted investigations into the possibilities for automated risk classification in 3D mammographic images. *Analysis of Parenchymal Texture with Digital Breast Tomosynthesis: Comparison with Digital Mammography and Implications for Cancer Risk Assessment* by Kontos *et al.* [60] identified texture features in both 2D digital mammograms and 3D tomosynthesis image volumes and performed linear regression analysis to determine associations between texture features and breast density.

The study used digital mammograms and breast tomosynthesis volumes from 71 women and found that correlations between texture features and percent density were stronger in 3D tomosynthesis volumes than in 2D mammograms, suggesting that information detected in 3D tomosynthesis volumes may provide a more accurate risk estimation [60]. The principal limitation of the study is the very small dataset used, which makes firm conclusions difficult. A further study using a larger dataset would allow for more reliable conclusions to be drawn, however large datasets of tomosynthesis images are not yet readily available due to

tomosynthesis being a new technique that is not yet in widespread clinical use.

2.3 Line & Edge Detection in Mammographic Images

Automated detection and classification of parenchymal structures within the breast has received a significant amount of research, although much of this is aimed at computer–aided diagnosis (CAD). A CAD system is an automated system that uses image analysis techniques to detect and classify malignant abnormalities in a screening mammogram with the intention of reducing false–negative and false–positive diagnoses [61]. The principles used to detect parenchymal structures for CAD could be applied to detect linear structures for risk assessment.

Linear structures in mammographic images: detection and classification, by Zwiggelaar *et al.* [62], provides a review of four different methods for detecting linear structures: Line Operator, Orientated Bins, Gaussian Derivatives and Ridge Detector. The paper made a comparison between the four methods with the aim of finding the most suitable for use with mammographic images.

Line Operator: Described by Dixon and Taylor [63] and applied to mammograms by Parr et al. [64], the Line Operator involves calculating the average grey–level, L, of a line of pixels centred on the target pixel at multiple orientations. The orientation with the largest L is chosen and the line strength, S, is given by $S = (L - N)$, where N is the average grey–level of a similarly orientated square of pixels centred on the target pixel. Grey–level values at inter–pixel points are obtained using bi–linear interpolation.

Orientated Bins: Described by Zwiggelaar *et al.* [65], the Orientated Bins method uses a circular region centered on the target pixel, which is divided into equal sectors. Pixels are assigned to the sector containing its centre. The average grey–level,

L, is calculated for each pair of opposing sectors. The orientation that maximises L is selected and the line strength, S, is given by $S = (L - N)$, where N is the average grey–level for the entire circular region. Orientated Bins has the advantage in that all orientations can be calculated in a single step, making it considerably less computationally intensive than the Line Operator. *Gaussian Derivatives:* Karssemeijer [66] described a method of line detection based on Gaussian derivatives [67, 65, 68]. The image is convolved with Gaussian second derivative kernels in three different orientations and the results, W, are combined arithmetically. The line strength, $W(\theta)$, and orientation, θ are given by

$$\begin{aligned} W(\theta) = & \frac{1}{3}\left(1 + 2\cos(2\theta)\right) W_0 \\ & + \frac{1}{3}\left(1 - \cos(2\theta) + \sqrt{3}\sin(2\theta)\right) W_{\frac{\pi}{3}} \\ & + \frac{1}{3}\left(1 - \cos(2\theta) - \sqrt{3}\sin(2\theta)\right) W_{\frac{2\pi}{3}} \end{aligned} \tag{2.1}$$

$$\theta = \frac{1}{2}\left[\arctan\left(\frac{\sqrt{3}\left(W_{\frac{\pi}{3}} - W_{\frac{2\pi}{3}}\right)}{W_{\frac{\pi}{3}} + W_{\frac{2\pi}{3}} - 2W_0}\right) + \frac{\pi}{2}\right]. \tag{2.2}$$

Ridge Detector: Lindeberg [69, 70] described a method based on second–order derivatives. The author proposed several measures of ridge strength, the one selected for analysis was the $A_{\gamma-norm}L$ index since this is designed to reduce the response to blobs. The line strength and orientation, θ are given by

$$A_{\gamma-norm}L = t^{2\gamma}\left(\left(L_{xx} - L_{yy}\right)^2 + 4_{xy}^2\right) \tag{2.3}$$

$$\cos(\theta) = \sqrt{\frac{1}{2}\left(1 + \frac{L_{xx} - L_{yy}}{\sqrt{\left(L_{xx} - L_{yy}\right)^2 + 4L_{xy}^2}}\right)} \tag{2.4}$$

where t is the scale, γ is a normalisation constant and L_{xx}, L_{yy} and L_{xy} are the second–order derivatives with respect to the subscripts.

At various stages of the experiment, both synthetic and real mammographic images were used, however, during the initial comparison of the various line detec-

tion methods only synthetic data was used. Using synthetic data has the advantage that the linear structures created on them have known lengths and positions, so the performance of the detection methods can be accurately measured [71, 72, 73]. With real images, however, the exact attributes of the linear structures cannot be determined to a high degree of accuracy, since the image must be captured correctly and interpreted by the radiologist, and different radiologists are likely to interpret the images slightly differently [74]. A disadvantage of synthetic images is that it is uncertain how accurately they represent real data. This is a problem that is present in all similar work on measuring the accuracy of line detection methods on mammographic images; one can never accurately measure the performance of a particular method against real mammographic data.

The four line detection methods were compared in terms of their signal-to-noise performance, orientation accuracy, width accuracy, and localisation accuracy. The signal-to-noise performance was determined using receiver operating characteristic (ROC) analysis and gives a measure of how accurate a detection method is at determining whether pixels are part of a linear structure or not. The area under the ROC curve indicates the degree of accuracy with the line operator method achieving 0.943, orientated bins achieving 0.916, Gaussian derivatives achieving 0.904, and ridge detector achieving 0.820. These results show that the line operator method has the best signal-to-noise performance, with the ridge detector having the worst performance.

Orientation and width accuracy were measured by comparing the orientation and width at each pixel with the known values from the original image. The cumulative angular error was calculated for each method and summarised as the standard deviation of angular error in order to show the orientation accuracy, and the width accuracy was presented as confusion matrices and summarised as a percentage correct value for each of the four methods.

Orientation accuracy results showed that the orientated bins method provided the best orientation accuracy (7.2° standard deviation), with line operator (11.8°), Gaussian derivatives (12.5°) and ridge detector (13.3°) all achieving slightly worse, but still reasonable results. Width accuracy measurements showed that the line operator method was the most accurate (73.4%), followed by orientated bins (68.2%), ridge detector (61.4%) and Gaussian derivatives being the least accurate (58.8%).

Before the localisation accuracy could be measured, the linear structures were extracted from the line strength and orientation images. This technique used to extract the linear structures first used non-maximal suppression (NMS) [75]. This involved comparing each pixel with those either side of it in a line perpendicular to the orientation direction and selecting pixels whose line strength was greater or equal to that of its neighbours. After removing those with a line strength below a threshold level, binary thinning [75] was then used and marked pixels were categorised according to their number of marked neighbours. Those with one marked neighbour were considered to be endpoints of linear structures, those with two marked neighbours were considered to be a part of a linear structure, whilst those with more than two marked neighbours were unmarked until none remained. The connecting paths of pixels were then extracted. The experiments in the paper ignored lines of less than 8 pixels in length.

Once the linear structures had been extracted, the localisation accuracy was measured for each of the four methods by measuring the distance from each pixel on linear structures on the known original image with the nearest point on an extracted linear structure, and calculating the cumulative distributions of localisation error. The results showed that the line operator method was again the most accurate, showing detected linear structures within two pixels of 86% of known linear structure pixels. The orientated bins method scored the second highest results (77%), with Gaussian derivatives coming third (64%) and the ridge detector method performing comparatively poorly (45%). The results showed quite a

significant difference in accuracy between the best and worst performing methods.

Throughout the experiments the line operator method has performed consistently well (coming first in three experiments and second in the remaining one), with the orientated bins method also performing quite reasonably (coming first in one experiment and second in the remaining three). Conversely, the ridge detector method has performed consistently poorly in comparison with other methods (coming last in three experiments and third in the remaining one), although in some of the experiments the difference between the results for different methods was not substantial.

The paper continues with further experiments using the line operator method on real mammographic images, and then attempting to classify the extracted linear structures to identify the type of structure (either edges, spicules, vessels, ducts, fibrous tissue, skin folds or other). The linear structures on the real mammographic images had been annotated by a radiologist, and then the extracted linear structures were classified based on their cross-sectional profiles and the results compared with the radiologist's annotations.

The benefit of knowing your linear structures in mammographic images [75] also by Zwiggelaar, *et al.* followed many similar procedures as [62], and aimed to investigate how the detection and classification of linear structures in mammographic images can improve the detection rate of microcalcifications and how it can also help to determine the malignancy of the abnormalities.

This paper recognised that certain abnormalities are associated with certain types of linear structures, particularly ducts and spicules, and that by detecting and classifying the linear structures, the detection rate of abnormalities can be improved. This experiment used twenty real mammographic images, half normal and half containing microcalcifications.

The classification of the extracted linear structures resulted in each linear structures being given a duct probability value, which can be used to help determine the malignancy of microcalcifications. As in the analysis above, ROC curves were produced to show firstly a comparison between the detection of microcalcifications based solely on the grey-level values of the original image, and the detection when based on these values multiplied by the line strength probability. The area under the curve, A_z, providing a representation of the detectability of the microcalcifications, increased from 0.74 to 0.93 when the linear structure information was taken into account. This showed a very clear improvement in the ability to detect microcalcifications in the images.

The second analysis conducted in the paper was to plot an ROC curve showing the classification of the malignancy of microcalcifications based on the duct probability values of the linear structures. An A_z value of 0.69 was achieved, indicating a reasonable ability to determine whether a microcalcification is malignant or benign based on duct probability.

Zwiggelaar *et al.* [76] demonstrate the detection and classification of linear structures based on statistical modelling of scale-orientation signatures. The paper stated that previous classification attempts [77] based on extracted scale information had not provided particularly good results and had widely differed from radiologists' annotations. An experiment was conducted using 29 real mammographic images, which were processed using the Line Operator to detect linear structures. These detailed structures were then annotated by an expert radiologist into several classes of structure (ducts, edges, fibrous tissue, skin folds, spicules, vessels and others). Orientation signatures were then extracted from those structures.

Results showed that the ROC analysis for the classification of spicules based on the signature approach, compared with that based on the cross-sectional profiles indicated that the new approach performed slightly worse at high sensitivity but

better at a lower sensitivity. Zwiggelaar *et al.* [76] continued to produce classification probability images using the same approach, in an attempt to show the spicule probability at pixel level, and then combined this information with the detection probability images to produce spicule images. These spicule images were accessed only visually, no quantitative results were provided to assess their accuracy. The authors indicated that there does appear to be a correlation between the resultant spicule images and the spicules in the original annotated images, but it was not perfect, since some linear structures not annotated as spicules were detected as possible spicules. The authors also noted that not all spicules have been annotated, which is another example of the previously mentioned problem of using radiologists' annotations as 'correct' results.

Cerneaz and Brady [78] described a method for detecting curvilinear structures in mammograms using a model–based ridge finder. The authors state that conventional ridge detection methods often perform poorly on mammographic images when linear structures can lie adjacent and closely parallel to each other. This causes conventional methods of reducing noise, such as smoothing, to severely reduce the effectiveness of the line detection. In order to overcome this, the authors use a statistical model to identify the set of pixels belonging to the linear structure components. This set is then expanded to fill small holes in the set and these regions are thinned to produce thinly–connected skeletons. The skeleton structures are separated at each junction into distinct branches and the set of branches is scanned for branches that can be merged or discarded due to their similarity. The branches are reconstructed to form a skeletal representation of the linear structure network. The results closely matched the linear structures observed by radiologists. This approach differed from many other approaches in that it focussed on detecting the entire network rather than any pixel–level measures of line strength and orientation.

In summary, a variety of methods for detecting linear structures in mammo-

graphic images have been demonstrated. Most of these methods have adopted a pixel–level approach, whereby each pixel is annotated with a measure of line strength and orientation [62], however others have used model–based approaches to detect the network of structures [78]. The pixel–level information could be useful for tasks such as the classification of structures, and measures of line strength for all pixels could be useful in risk assessment, where the density of linear structures has been linked to mammographic risk [22]. Of the pixel–level methods reviewed, the Line Operator has been shown to outperform several other methods [62]. The problem of detecting linear structures in three–dimensional breast images has not been well–studied, and there is little literature available on this topic, however it may be possible to adapt an existing method to work in three dimensions.

2.4 Mammographic Density Estimation

Conventional classification of mammographic density by observers has several limitations. The task is time–consuming for the readers, and the inevitable inter– and intra–observer variabilities lead to inconsistent and unreliable results [16, 79]. These limitations led to the development of semi–automated and automated methods as a way of quantifying breast density more reliably and consistently.

Some methods, such as interactive thresholding described by Byng, *et al.* [16], are semi–automated. The interactive thresholding method requires the reader to select two thresholds. The first threshold is to outline the breast area whilst the second threshold best outlines areas of dense tissue. Both inter– and intra–observer variability were shown to be low using this method. However it remained present as a limitation since the method remains dependent on the judgement of the observer [16, 80].

Automatic segmentation of mammographic density by Sivaramakrishna *et al.* [81] described an entirely automated method for segmenting dense tissue. The authors firstly calculated a variation image which was intended to suppress small bright areas (such as vessels, noise, etc.) whilst maintaining large bright areas (such as dense tissue). Secondly, a threshold was determined automatically using the *Minimum Error Thresholding* technique [82] and applied in order to delineate the dense tissue areas. The method was tested using a small set of 32 images; the results showed that the density measurements correlated closely with the measurements of mammographers. The difference between results obtained using the algorithm and those of mammographers was similar to the differences in results between mammographers.

A further study using a larger data set would provide further information as to the accuracy of measurements obtained using the method, since it is difficult to draw firm conclusions from such a small trial. However, a fully–automated method such as this offers the potential for significant benefits as it would overcome observer variability problems inherent in systems that depend on the judgement of an observer.

Zhou *et al.* [83] described an automated method along similar principles to Sivaramakrishna *et al.*. The breast area is subjected to an adaptive dynamic range compression technique designed to reduce the grey–level intensity in background tissue whilst enhancing larger bright areas. A threshold is then determined automatically using the grey–level histogram in order to segment the dense tissue.

Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort by Klifa *et al.* [84] used a semi–automated segmentation method to estimate breast density in 3D MRI image volumes. The method involved a computer system drawing a contour around the dense tissue areas using a combination of Bezier splines and a Laplacian of Gaussian filter. The user could influence the

contour by placing control points of Bezier splines close to the edges of the region of interest.

The authors performed a study to compare the performance of the method using 2D mammograms and 3D MRI images. Mammograms from a set of 35 women were assigned percent density measures using a predominantly manual method, whereby a reader delineated the area of breast density by hand. A computer program then analysed these to calculate percent density measures, which were then compared with the measures obtained from the MRI volumes. The overall results showed a moderate correlation between density estimations from 2D and 3D methods, however when divided into low–risk (BIRADS classes 1 and 2) and high–risk (BIRADS classes 3 and 4) groups the low–risk group showed much stronger correlation between 2D and 3D measures, whereas the high–risk group showed poor correlation. The authors suggest that this may indicate that 2D density estimation is less accurate for more dense breasts, and that therefore estimates using 3D image volumes may provide a more accurate risk estimation for this group [84]. Limitations of the study include the very small dataset of only 35 women, and that different methods were used to assess density in 2D and 3D images. The small dataset makes any firm conclusions difficult to establish, whilst differences in the estimation methods could easily account for much or all of the observed differences. As previously stated, however, large datasets of tomosynthesis images are not yet readily available due to tomosynthesis being a new system that has only recently gained approval for clinical use.

Yaffe [85] provides an overview of various approaches to measuring breast density. He states that initial methods of estimating breast density were entirely subjective, however in recent years more objective and qualitative approaches have been developed. The most straightforward method, known as planimetry, involved manually tracing around areas of dense tissue on a screen film mammogram using an instrument called a planimeter. This provides the total area of dense tissue. A

similar measurement of the total breast area is obtained and the first measurement is divided by the second to obtain the proportion of breast area considered dense. Whilst this method is simple to perform, it is highly labour–intensive and prone to inter– and intra–observer variability.

Semi–automated methods such as interactive thresholding [16] provide a less time–consuming alternative to planimetry when using digital or digitised mammograms. These methods have been used successfully in many clinical studies [85], and generally involve the user interactively selecting a threshold that as closely as possible segments the dense from fatty tissue, and often a second threshold that segments the breast area [85, 16]. Yaffe argues that a principle shortcoming of these techniques is that a single threshold that clearly segments all dense tissue from fatty tissue is unlikely to exist, resulting in a degree of subjectivity in the selection of the threshold and consequently inconsistent results. This can be partially overcome by training and the adherence to reading standards, but is unlikely to be eliminated, giving rise to the desire for fully–automated methods.

Chang *et al.* [86] and Glide-Hurst *et al.* [87] have described fully automated methods based on thresholding, in addition to those by Sivaramakrishna *et al.* [81] and Zhou *et al.* [83]. These methods generally use apply a function to suppress background noise and enhance larger bright areas that would represent dense tissue, prior to the automated selection of a threshold by analysing the grey–level histogram. These methods represent an improvement over the semi–automated methods since they remove the degree of subjectivity, however they may not always correctly identify the most suitable threshold [85].

Yaffe states that there has been relatively little work on comparing various methods for estimating breast density. Martin *et al.* [88] compared density measurements derived from several qualitative, quantitative and semi-automated methods and found large differences between assessments based on qualitative

and quantitative methods. The authors found that qualitative methods were less reproducible and tended to over–estimate the proportion of breast density.

In summary, fully automated methods to segment the dense areas of breast tissue have been developed [81, 83, 86, 87] and such methods are expected to produce the most reproducible, objective and qualitative estimates of breast density [85], since they remove user involvement in the segmentation process.

2.5 Mammographic Risk Classification

Various studies have been published investigating mammographic risk classification based on the analysis of mammographic images. *Texture Based Mammogram Classification and Segmentation* by Gong *et al.* [89] demonstrated the use of a texture–based method to classify mammograms into Wolfe classes. Textons were identified by extracting for each pixel in each of a training set of 200 images, the greyscale levels of the pixels in an $N \times N$ square around the pixel. These greyscale levels were then row reordered to form 'image patch' vectors. These were used as the initialisation for a Hidden Markov Random Field [90]. The resulting vectors were then aggregated over images from the same Wolfe class and clustered together, with textons identified using K–means clustering.

Five textons were chosen from each class for use in the classifier. For each test image, each pixel is labelled with the closest texton in the feature space, and a histogram is constructed of these texton labellings. These were then compared with the models learnt using the training set and assigned to the closest match. Final results showed correct classification rates of 92%, 81%, 83% and 95% for Wolfe classes 1–4 respectively [89]. The test set consisted of only 43 images and the authors do not provide a breakdown of the number of images belonging to each Wolfe classification.

Petroudi *et al.* [91] described an approach that was also based on texture classification using a texton dictionary developed from a training set. Each mammogram in the training set is firstly segmented to remove the background and pectoral muscle, then filtered using Maximum Response 8 filter bank [92]. The filter responses are clustered over all images per BIRADS class using the K–means algorithm and the cluster centres represent the textons. This results in a library of 10 textons per BIRADS class, 40 textons in total. Each pixel in each training mammogram is then assigned a label by the texton that lies closest to it in the filter response space, and a texton histogram is computed for each mammogram.

During classification, each test mammogram is segmented, filtered and mapped to a texton histogram using the same method. This histogram is compared to all training models and the nearest model is found using the χ^2 comparison by comparing its texton distribution to the learned models. The method achieved an overall classification accuracy of 76% when evaluated using the MIAS dataset [23].

Bovis and Singh [93] described an alternative approach to feature extraction and classification. Each mammogram is initially segmented to remove background area, and four groups of texture features are extracted in addition to a series of statistical features. The texture features include 1) extracting 15 features by constructing Spatial Grey Level Dependency (SGLD) matrices [94], 2) the total spectral energy from 10 equidistant analysing rings from the power spectrum following the application of the Fourier transform, 3) the total texture energy for this mask combination by convolving the image with each combination of Laws' texture masks, and 4) four features (standard deviation, mean, skewness and kurtosis) following application of the Discrete Wavelet Transform. Classification in to BIRADS classes was conducted using an Artificial Neural Network. Evaluation using 377 images from the DDSM dataset [95] using a 10–fold cross–validation methodology produced an overall classification accuracy of 71.4%.

Oliver *et al.* [96, 97] presented a texture–based approach to risk classification by using texture information to segment dense from fatty tissue areas. Their investigation involved using a clustering algorithm to group pixels of similar grey–level intensities together, ultimately resulting in just two clusters of pixels representing dense and fatty tissue types.

Two different methods, Local Binary Patterns [98] and co–occurrence matrices, were used to extract features from the segmented images for use in a classifier. During the experiment 322 images were classified into BIRADS classes using a leave–one–out classification methodology. The experiment investigated and compared three different clustering algorithms (Fuzzy C–Means, Normalised Cuts and the Mean Shift algorithm), the two different feature extraction methods and four different classifiers (K–Nearest–Neighbour with Sequential Forward Selection (SFS) feature selection, Linear Discriminant Analysis, $C4.5$ Decision Tree and Support Vector Machine with a polynomial kernel).

Results for the comparison of clustering methods showed 65%, 78%, 73% and 77% correct classification when no segmentation, Fuzzy C–Means, Normalised Cuts and the Mean Shift algorithm were used respectively, along with the Local Binary Patterns feature extraction and the K–Nearest–Neighbour classifier. The result for no segmentation showed that using a clustering algorithm improves classification accuracy significantly. The performance of the C–Means and Mean Shift algorithms were similar, whilst the Normalised Cuts algorithm performed less–well.

The comparison between feature extraction methods was largely inconclusive, showing 78% and 79% correct classification for the Local Binary Patterns and co–occurrence matrices methods respectively, when used with the Fuzzy C–Means clustering and K–Nearest–Neighbour classifier. Meanwhile, the results for the comparison of classifiers showed 78%, 65%, 68% and 65% correct classification

for the K–Nearest–Neighbour, Linear Discriminant Analysis, $C4.5$ Decision Tree and Support Vector Machine classifiers respectively, when used with the Fuzzy C–Means clustering and Local Binary Patterns feature extraction.

This result showed that the K–Nearest–Neighbour classifier outperformed all other classifiers significantly. The authors state that this is indicative of the difficult task of finding a true boundary between classes, and so it is unsurprising that a linear classifier such as Linear Discriminant Analysis performs poorly [96]. The relatively poor performance of the Support Vector Machine classifier is perhaps a surprising result and may indicate mis–configuration. The Support Vector Machine classifier used a 10–fold cross–validation methodology instead of leave–one–out, resulting in a smaller training set than for other classifiers, which may also have influenced the results.

Different approaches have been developed to automatically classify mammograms in to risk categories, although since these studies are often conducted using different datasets and different risk metrics it is not always possible to draw firm conclusions as to the relative performance of the different methods. Classification methods commonly comprise two distinct steps: firstly the extraction of features from the image, and secondly the use of these features in a classifier to assign the image to a particular risk category. Some studies used texture analysis based on textons to segment the image [89, 91]. These methods used texton labels to construct a texton histogram for each image and classification is conducted by comparing the histogram of a test image against each histogram in a training set. Other studies extracted texture and statistical features from the image for use in a classifier [93, 96].

Work by Oliver *et al.* [96] evaluated a range of segmentation methods, feature extraction methods and classifiers. Their approach was to use a clustering method to segment the image in to dense and fatty tissue types, followed by a

method to extract features for use in a classifier, and finally classification. The evaluation included both linear and non–linear classifiers and concluded that, with the exception of the K–Nearest–Neighbour classifier, there was little difference in the performance of the other three classifiers evaluated suggesting that classes are often not well–separated and finding a boundary between classes is not straightforward.

Chapter 3

2D Risk Assessment

3.1 Introduction

The purpose of this chapter is to investigate the relationship between the density of linear structures and mammographic risk. Initial work involves the selection and implementation of a line detection methodology and experiments to investigate whether statistically significant differences in linear density can be detected in mammograms of different risk categories. This would suggest that linear density may be used as a predictor for mammographic risk. Further experiments use linear density in a classifier to automatically place mammograms in to risk categories and compare against a conventional density–based classifier.

This chapter focusses on the use of conventional two–dimensional mammograms in order to establish any relationship between linear density and mammographic risk, since this technology is in widespread use and large datasets are readily available for research purposes.

3.2 Data

Experiments are conducted using images from the Mammographic Image Analysis Society (MIAS) database [23]. This database is composed of left and right mammograms from 161 women taken from the screening population. The images are digitised into $50\mu m \times 50\mu m$ pixels. One image (mdb295) had become corrupted in the copy available, resulting in a set of 321 images that are used throughout this chapter.

Three classification models are used: Tabár patterns [22], Boyd SCC classes [28] and Breast Imaging Reporting and Data System (BIRADS) classes [27]. Tabár's classification consists of five *patterns*, where patterns I–III represent a low risk of developing breast cancer, and patterns IV–V indicate a higher risk. Screening tests have shown that cancer prevalence in women with patterns IV-V is approximately twice that in women with patterns I-III [22]. The Boyd SCC model consists of a scale of six classes where class 1 indicates the lowest risk and class 6 indicates the highest risk. Finally, the BIRADS classification uses a scale of four classes, where class I represents a low risk and class IV represents a high risk.

The 321 mammograms were classified by an expert radiologist according to each classification model. Furthermore, two additional expert radiologists classified each mammogram according to the BIRADS risk classification metric. Consensus between the classifications of the three individual radiologists was used, as is common in screening mammography, in order to improve the reliability of the classifications.

Where disagreement occurred between the experts, the consensus was determined using the following strategy: where two out of three radiologists agree, the majority value was used, and where all three experts disagreed, the median value was used. Of the 321 mammograms, all three experts agreed in 138 cases, two

experts agreed whilst a third disagreed in 171 cases, and in 12 cases all three experts disagreed.

Tabár Pattern	**I**	**II**	**III**	**IV**	**V**
Images	119	53	40	81	28

Table 3.1: The distribution of images according to the Tabár classification metric.

Boyd SCC Class	**1**	**2**	**3**	**4**	**5**	**6**
Images	6	60	46	75	90	44

Table 3.2: The distribution of images according to the Boyd classification metric.

BIRADS Class	**I**	**II**	**III**	**IV**
Radiologist 1	129	79	69	44
Radiologist 2	86	112	80	43
Radiologist 3	59	86	142	34
Consensus	87	103	94	37

Table 3.3: The distribution of images according to the BIRADS classification metric for each radiologist and the consensus.

The distribution of images in the dataset across the three risk classification metrics can be seen in Tabs. 3.1–3.3. The tables show that a range of mammograms of all risk levels is present in the dataset. Amongst the 321 mammograms in the dataset, 115 contained one or more abnormalities.

Figures 3.1–3.3 provide an illustration of some typical mammograms covering a range of risk levels from the MIAS database. The images have been masked to remove the pectoral muscle and the non–breast area.

3.3 Linear Structure Detection

3.3.1 Method

The objective of investigating the relationship between the density of linear structures and mammographic risk has necessitated the development of a robust method

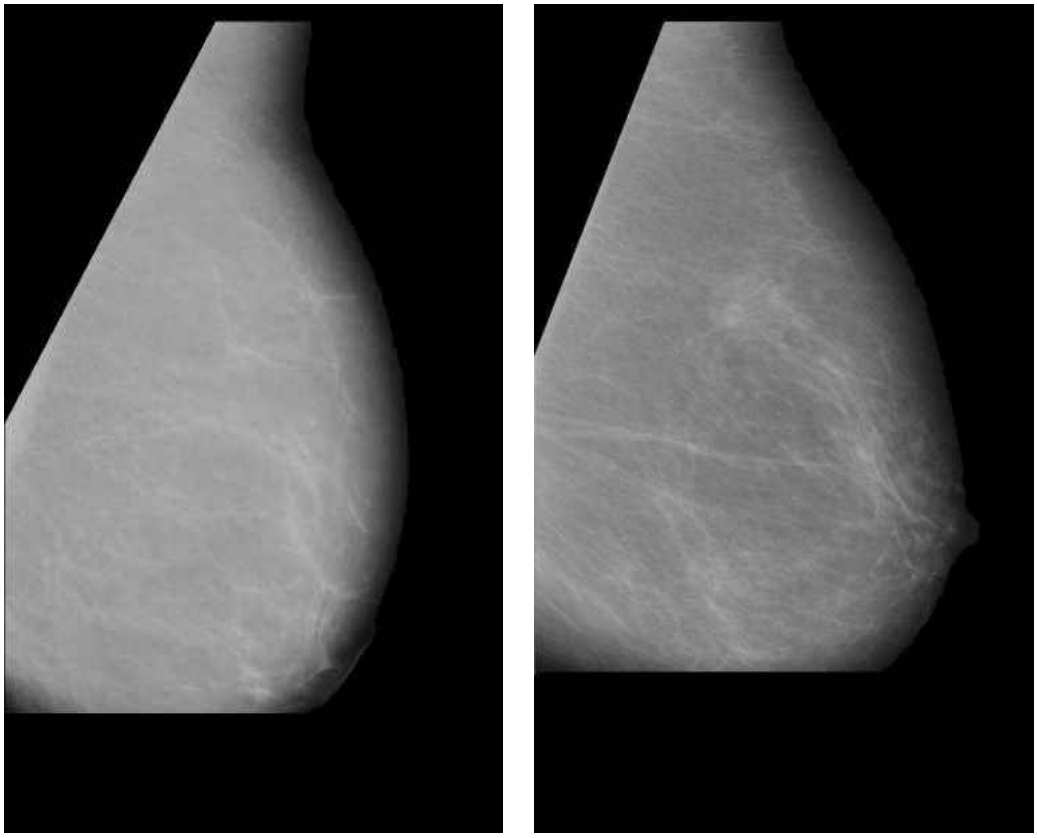


Figure 3.1: Typical low risk mammograms from the MIAS database. Left image shows a mammogram of Tabár pattern II, Boyd SCC class 1, BIRADS class I. Right image shows a mammogram of Tabár pattern II, Boyd SCC class 2, BIRADS class I.

for detecting linear structures in mammographic images. A study of various methods for detecting linear structures in mammograms [62] showed that Dixon and Taylor's line operator [63] is more accurate than other methods. As such, the line operator was used in our experiments. The method produces a measure of line strength and orientation for each pixel in an image.

The line orientation is determined by calculating the mean pixel brightness of a line of pixels running through the target pixel at a range of orientations, as shown in Fig. 3.4. The orientation with the largest mean brightness is taken to be the line orientation. The line strength, S, is then given by

$$S = (L - N), \tag{3.1}$$

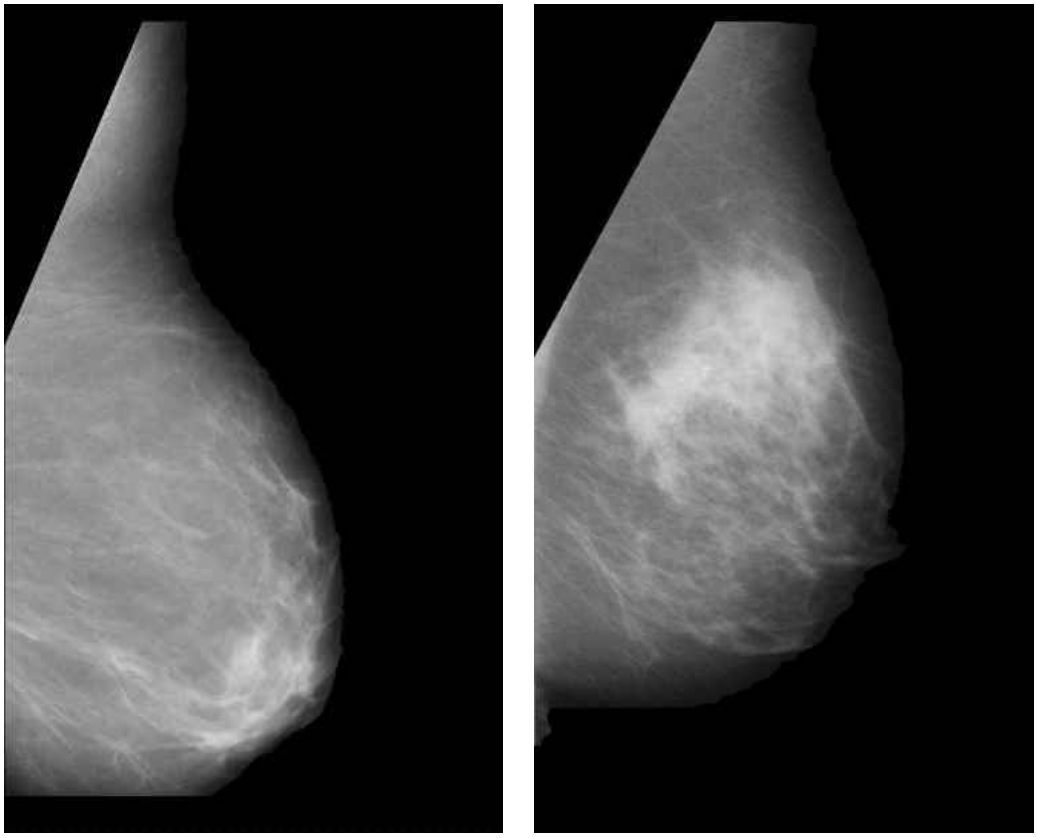


Figure 3.2: Typical moderate risk mammograms from the MIAS database. Left image shows a mammogram of Tabár pattern III, Boyd SCC class 3, BIRADS class II. Right image shows a mammogram of Tabár pattern I, Boyd SCC class 4, BIRADS class III.

where L is the mean brightness of the line of pixels, and N is the mean brightness of a similarly orientated square of pixels as shown in Fig. 3.5. A square is used rather than a circle to ensure that the background intensity for the full length of the line is measured to avoid a bias towards the centre–point of the line. Our experiment used a line length of five pixels and twelve orientations as suggested by earlier work [62]. Bilinear interpolation was used to calculate the pixel values where the required points fell between actual pixels.

The method makes the assumption that the lines under detection are approximately one pixel thick, since otherwise part of the line will be included in the 'background' measure resulting in a lower line strength measurement. In order to overcome this, a multi-scale approach was used in order to detect lines of a range of thicknesses and the resultant images were combined to produce line strength

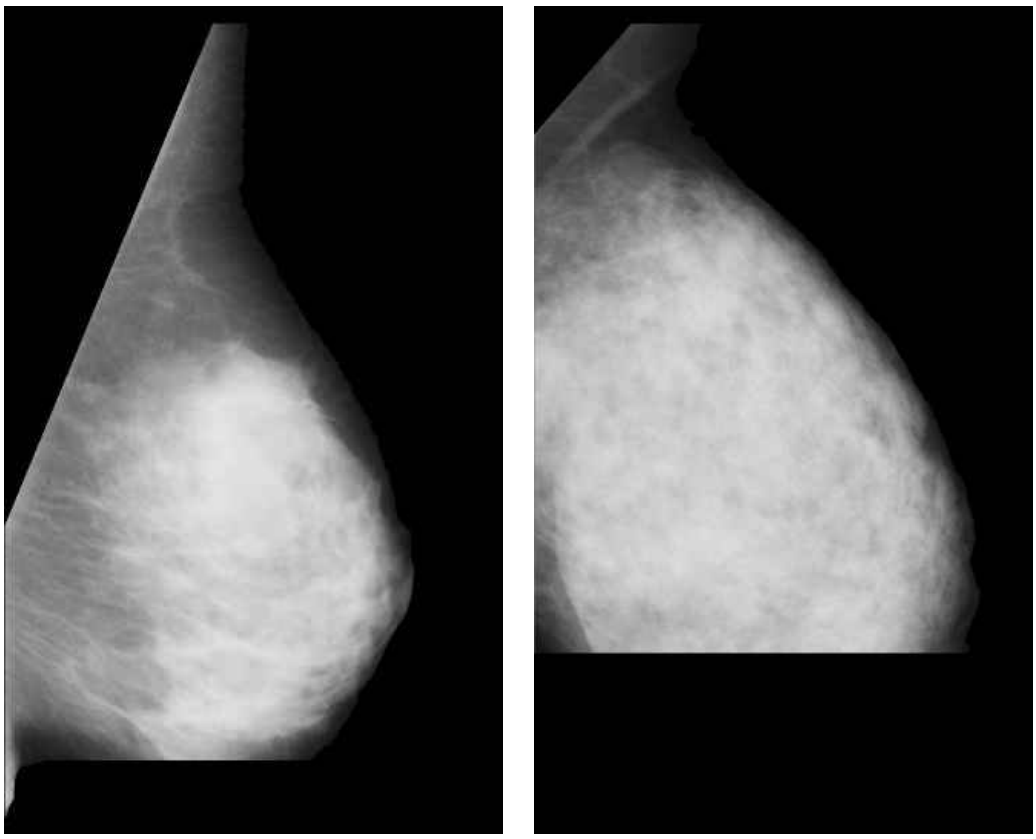


Figure 3.3: Typical high risk mammograms from the MIAS database. Left image shows a mammogram of Tabár pattern IV, Boyd SCC class 6, BIRADS class IV. Right image shows a mammogram of Tabár pattern V, Boyd SCC class 6, BIRADS class IV.

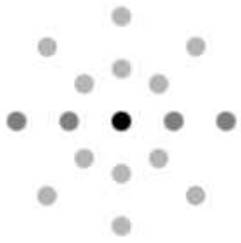


Figure 3.4: Diagram showing the line operator method selecting an orientation. The average grey–scale level is calculated for lines of pixels running through the target pixel at a range of orientations. The orientation with the greatest average line strength is selected.

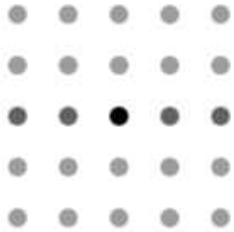


Figure 3.5: Diagram showing the line operator method measuring the line strength. The average grey–scale level is calculated for lines of pixels running through the target pixel, L, and for a similarly orientated square of pixels, N. The line strength, S, is given by Eq. 3.1.

values for pixels at the original scale. Scaling of the images was achieved firstly by blurring the image using a 3x3 Gaussian kernel and subsequently subsampling to provide a resultant image of half the width and height of the original. A Gaussian of 3x3 ($\sigma = 0.85$) was selected because it was felt that a larger Gaussian may blur the image too much and result in a loss of clarity which may lead to poorer results from the subsequent linear structure detection. The line strength results from each scale were combined by enlarging the scaled images to the size of the original and for each pixel taking the maximum line strength value from the set of line strength images. Our approach comprised processing with the line operator at three scales, since this appeared to produce the most reasonable output for the images under examination.

Finally, the pixel line strengths were thresholded to remove background texture, since this was the simplest method for obtaining a measure of linear density. Using a line length of 5, the measures of line strength in 8–bit images (with pixel values $0 - 255$) can reach a theoretical maximum of 204 since this is the highest value of S that can be obtained (see Eq. 3.1), however the results showed that most pixels had line strength values in the range $0 - 30$. A range of threshold values were chosen experimentally, and two values (4/204 and 6/204) were used

for our analysis as they removed most background noise whilst maintaining most of the linear structure information (see Figs. 3.6–3.8 (c) and (d)).

Experimental procedures to select thresholds involved performing the linear structure detection on a small set of 4 images selected randomly from the dataset. The detection was performed on these images using thresholds at integer increments between 1/204 and 30/204. Initial visual inspection resulted in all thresholds below 3 and above 8 to be discarded, since they clearly either removed too much linear structure information or not enough background information. The line detection was then performed on a larger set of 50 randomly selected images at each threshold in the remaining range and pairwise Mann-Whitney tests were performed similar to those shown in Sec. 3.3.2 in order to assess the suitability of each threshold. Based on these tests, the thresholds of 4/204 and 6/204 were selected for further analysis, since they provided a good level of distinction between the range of risk categories.

Figures 3.6–3.8 (b) show examples of low, moderate and high risk mammograms following processing with the line operator. Figures 3.6–3.8 (c) and (d) show examples of the resultant images after thresholding. Subsequently, the results were analysed for differences between images of each Tabár pattern, Boyd SCC class and BIRADS class.

3.3.2 Results

The relative linear structure density for the various Tabár patterns, Boyd SCC classes and BIRADS classes are shown in Figs 3.9, 3.10 and 3.11, respectively. These graphs provide an overview of the difference between the patterns and classes, and for more detailed analysis, Mann-Whitney U tests were performed on each pair of Tabár patterns, Boyd SCC and BIRADS classes. These are shown in Tabs 3.4, 3.5 and 3.6, respectively. Parametric tests, such as analysis of variance

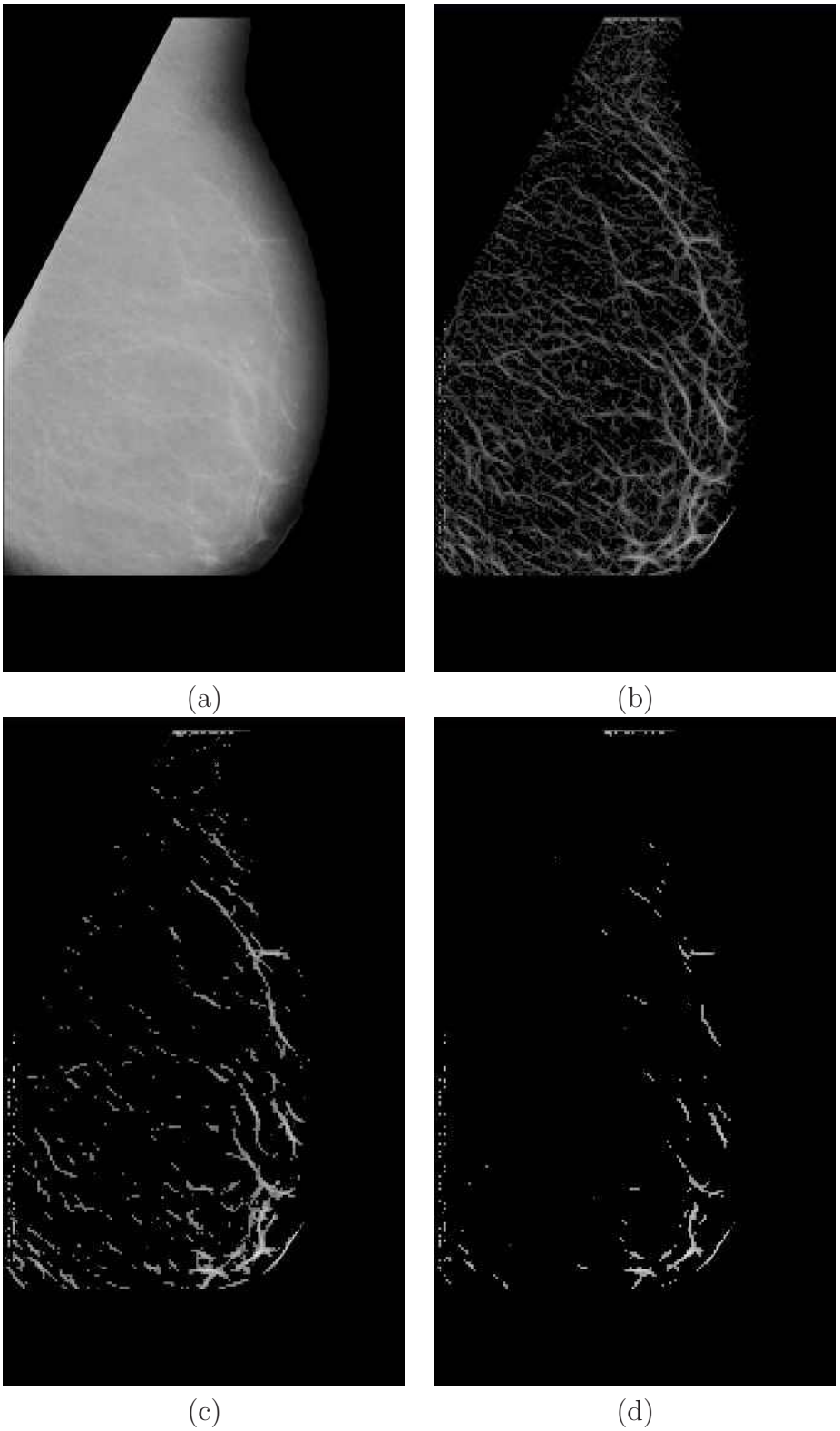
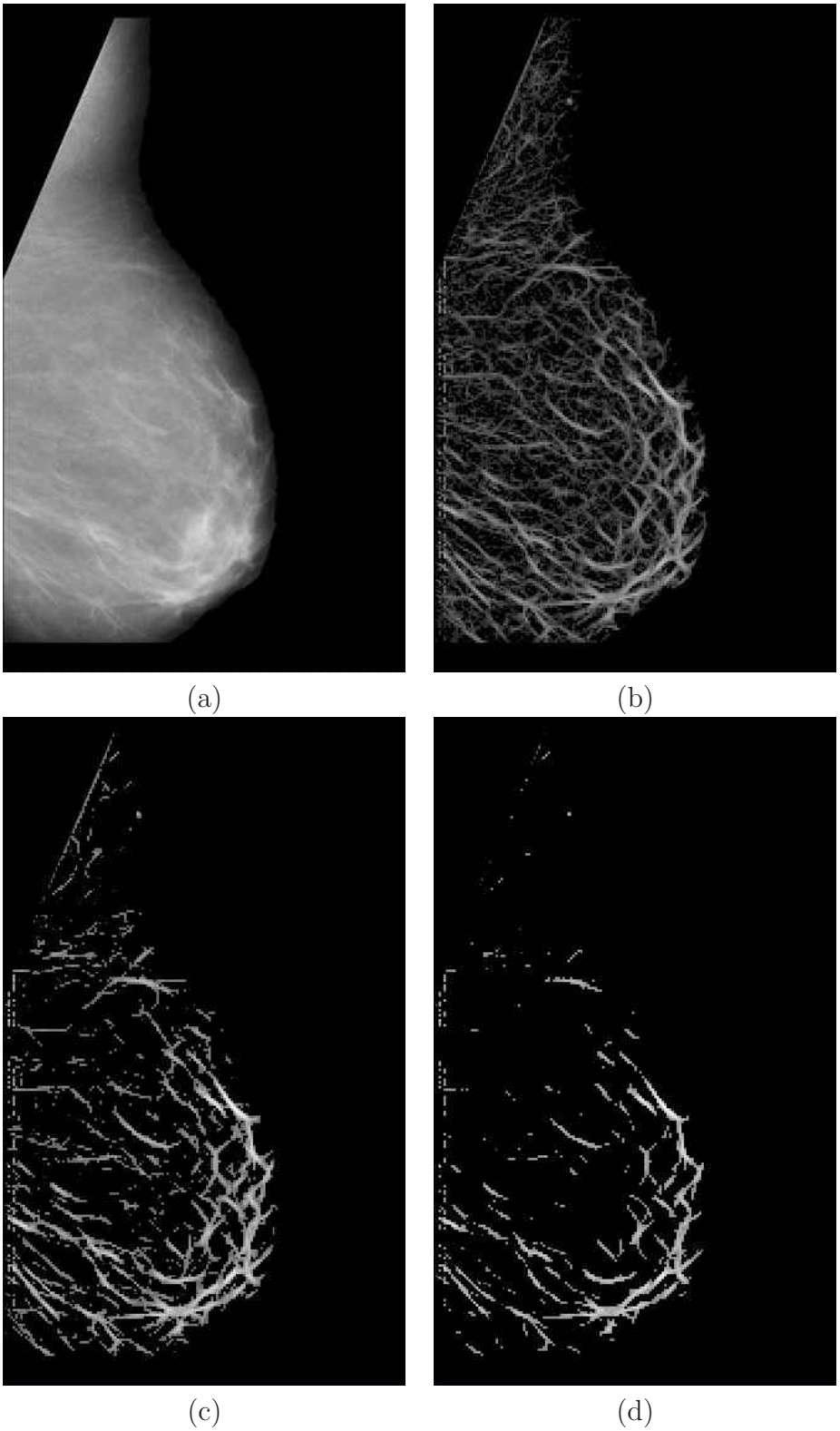


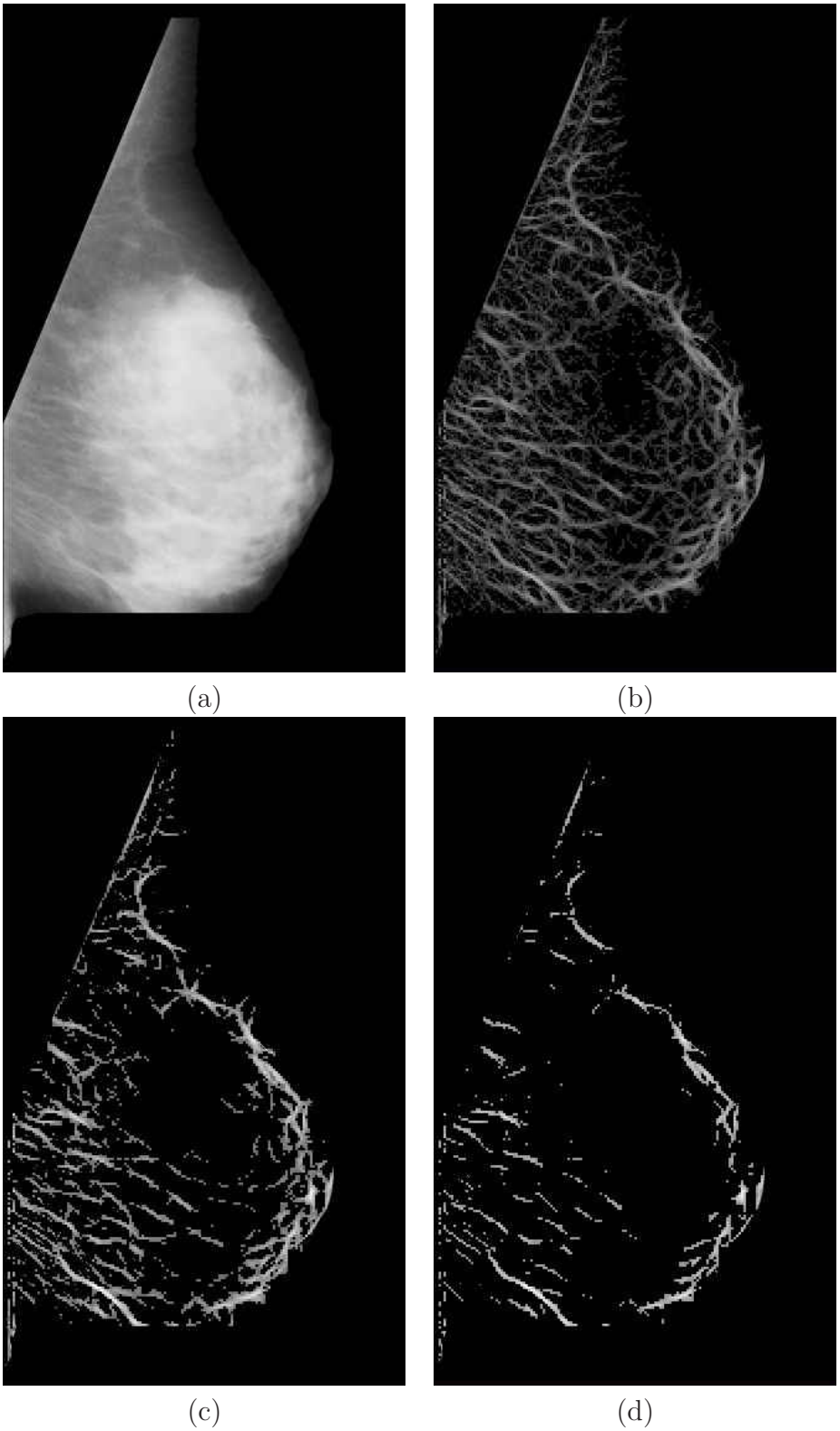
Figure 3.6: A typical mammogram of Boyd SCC class 1/Tabár pattern II/BIRADS class 1 (low risk).
(a) – original mammogram.
(b) – the results after processing with the line operator.
(c) and (d) – the results after thresholding at 4/204 and 6/204 respectively.
The lines have been enhanced for viewing.



(a) (b)

(c) (d)

Figure 3.7: A typical mammogram of Boyd SCC class 3/Tabár pattern III/BIRADS class 2 (moderate risk).
(a) – the original mammogram.
(b) – the results after processing with the line operator.
(c) and (d) – the results after thresholding at 4/204 and 6/204 respectively.
The lines have been enhanced for viewing.



(a) (b)

(c) (d)

Figure 3.8: A typical mammogram of Boyd SCC class 6/Tabár pattern IV/BIRADS class 3 (high risk).
(a) – the original mammogram.
(b) – the results after processing with the line operator.
(c) and (d) – the results after thresholding at 4/204 and 6/204 respectively.
The lines have been enhanced for viewing.

(ANOVA) tests were not used because the test data did not fulfil the necessary assumptions, however the relatively small number of classes made pairwise Mann-Whitney tests possible.

The Mann-Whitney test results provide an indication as to whether there was a statistically significant difference between two classes of images in our data set. A significant difference would suggest that it may be possible to reliably distinguish between different classes of images.

As mentioned, two threshold values were used in our analysis (4/204 and 6/204). This is to demonstrate the effects of varying the threshold level and thus including more or less of the linear structure information.

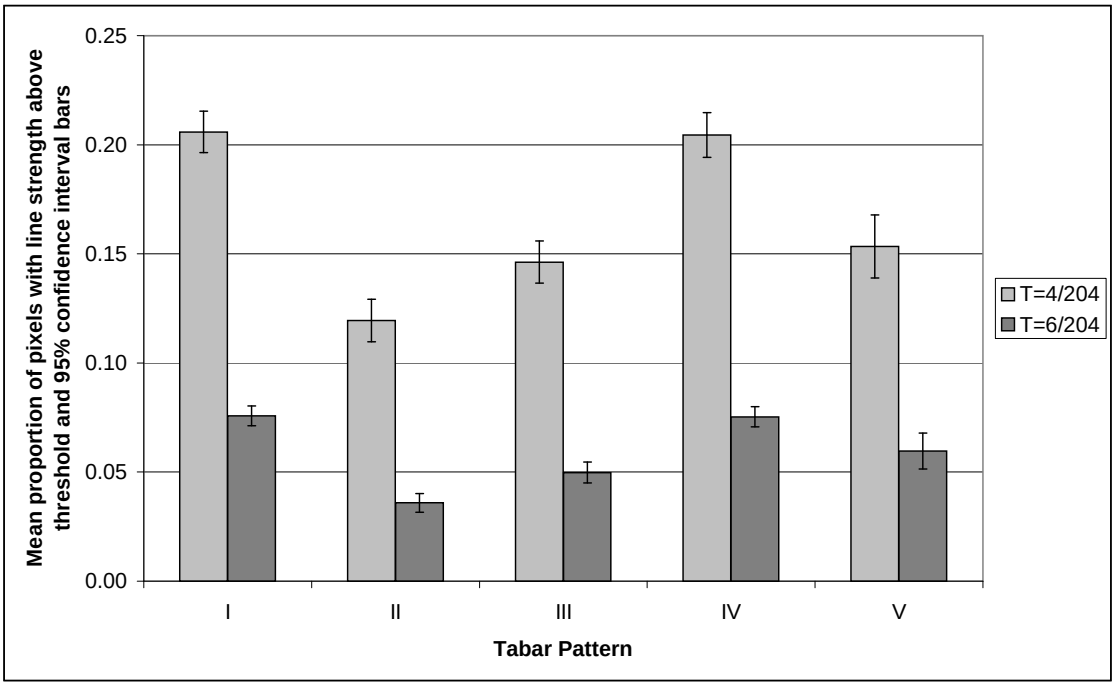


Figure 3.9: Graph showing the mean proportion of pixels with line strengths above each threshold T and 95% confidence intervals of the mean for images of each **Tabár** pattern.

The results of the analysis by Tabár pattern (see Fig. 3.9, Tab. 3.4) demonstrate the ability to reliably distinguish between patterns II–III (low risk) and patterns IV–V (high risk) at a threshold of 6/204, with low risk pattern I being indistinguishable from high risk pattern IV. The results at a threshold of 4/204

Table 3.4: The p-values obtained by Mann-Whitney tests on each combination of **Tabár** patterns for each threshold T. Results not significant at $\alpha = 0.05$ are shaded.

Tabár	**II**	**III**	**IV**	**V**
I	0.0000	0.0000	0.7938	0.0000
II		0.0003	0.0000	0.0002
III			0.0000	0.3597
IV				0.0000

$T = 4/204$

Tabár	**II**	**III**	**IV**	**V**
I	0.0000	0.0000	0.9069	0.0059
II		0.0001	0.0000	0.0000
III			0.0000	0.0336
IV				0.0050

$T = 6/204$

are less promising, since the low risk pattern III becomes indistinguishable from high risk pattern V.

The results of the analysis by Boyd SCC class (see Fig. 3.10, Tab. 3.5) differ somewhat, and it is clear that the proposed method is able to distinguish between classes 1–3 and classes 4–6 at both thresholds, with a general trend through classes 1–4/5 indicating that a greater linear density is indicative of a greater risk. The three lower risk classes are each distinguishable from all other classes, whilst the three higher risk classes are distinguishable from the three lower risk classes, but are indistinguishable from one another at $T = 6/204$. At $T = 4/204$ classes 4 and 6 become distinguishable from one another.

Results from the analysis by BIRADS class (see Fig. 3.11, Tab. 3.6) demonstrate the ability to distinguish the low risk class 1 from all other classes, and the higher risk class 3 from all other classes. However, it is not possible to distinguish the lower risk class 2 from the higher risk class 4. The trend shown here differs from the analysis by Boyd class; here the highest risk class 4 indicates a markedly lower average linear density than the moderately high risk class 3 and a similar density to the lower risk class 2. Whilst Boyd class 6 does show a slightly

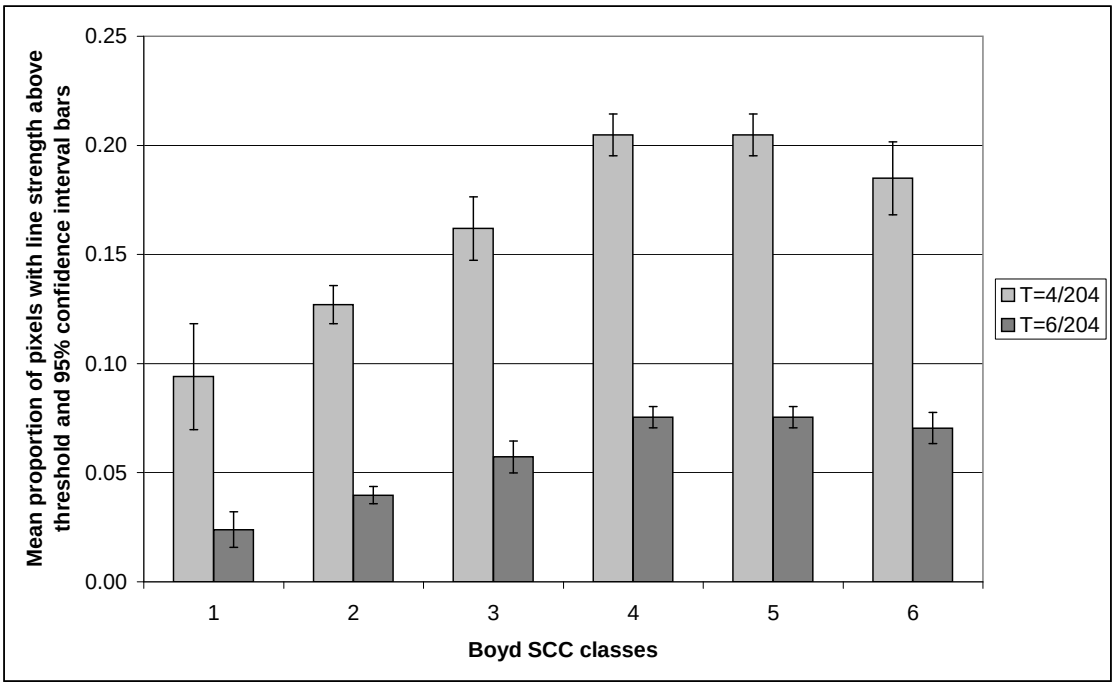


Figure 3.10: Graph showing the mean proportion of pixels with line strengths above each threshold T and 95% confidence intervals of the mean for images of each **Boyd SCC** class.

Table 3.5: The p-values obtained by Mann-Whitney tests on each combination of **Boyd SCC** classes at each threshold T. Results not significant at $\alpha = 0.05$ are shaded.

Boyd	**2**	**3**	**4**	**5**	**6**
1	0.0332	0.0023	0.0001	0.0001	0.0008
2		0.0002	0.0000	0.0000	0.0000
3			0.0000	0.0000	0.0362
4				0.6693	0.0287
5					0.0833

$T = 4/204$

Boyd	**2**	**3**	**4**	**5**	**6**
1	0.0155	0.0009	0.0001	0.0000	0.0004
2		0.0001	0.0000	0.0000	0.0000
3			0.0000	0.0000	0.0031
4				0.7621	0.4828
5					0.6612

$T = 6/204$

lower linear density than class 5, it is much less marked. This is reflected in the Mann-Whitney tests.

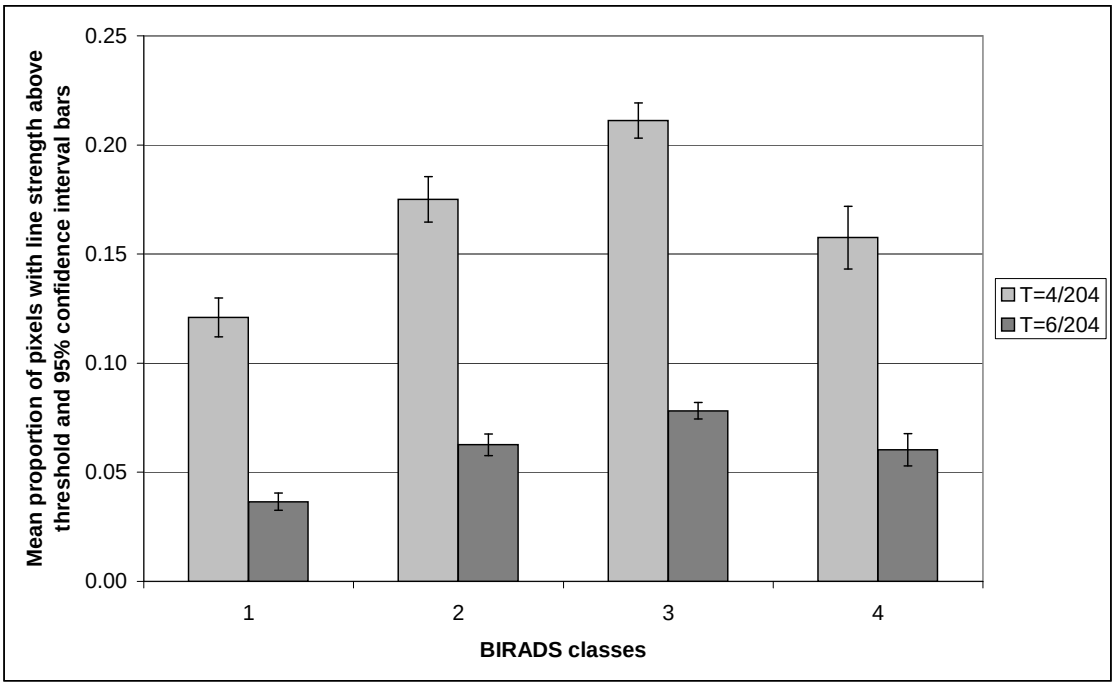


Figure 3.11: Graph showing the mean proportion of pixels with line strengths above each threshold T and 95% confidence intervals of the mean for images of each **BIRADS** class.

Table 3.6: The p-values obtained by Mann-Whitney tests on each combination of **BIRADS** classes at each threshold T. Results not significant at $\alpha = 0.05$ are shaded.

BIRADS	**2**	**3**	**4**
1	0.0000	0.0000	0.0000
2		0.0000	0.0771
3			0.0000

$T = 4/204$

BIRADS	**2**	**3**	**4**
1	0.0000	0.0000	0.0000
2		0.0000	0.8636
3			0.0002

$T = 6/204$

3.3.2.1 A Comparison of The Observed Results with Tabár's Expected Results

The risk assessment metric developed by Tabár is based on the relative proportions of four structural components. These are adipose tissue, fibroglandular tissue, nodular density and linear density. It is the proportion of the linear density component that is related to this work and Tabár's expected values for each pattern are shown in Fig. 3.12.

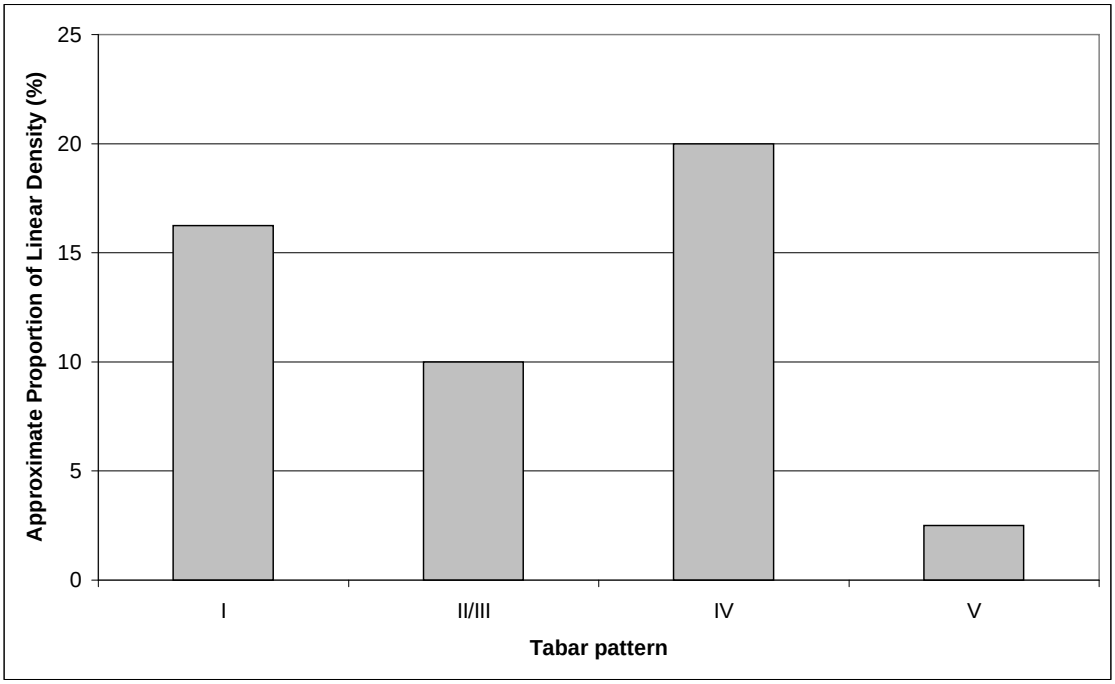


Figure 3.12: Graph showing Tabár's expected proportion of linear density in each pattern, taken from [22]. Note that due to their similarity, Tabár does not provide separate values for patterns II and III.

When the observed results at a threshold of 6/204 are normalised to Tabár's expected results (see Fig. 3.13), we can see that the results for patterns I-II are reasonably good approximations of the expected values, whilst the result for pattern IV under–represents Tabár's expected result. The result for pattern V is an anomaly which is discussed in Sec. 3.3.3.

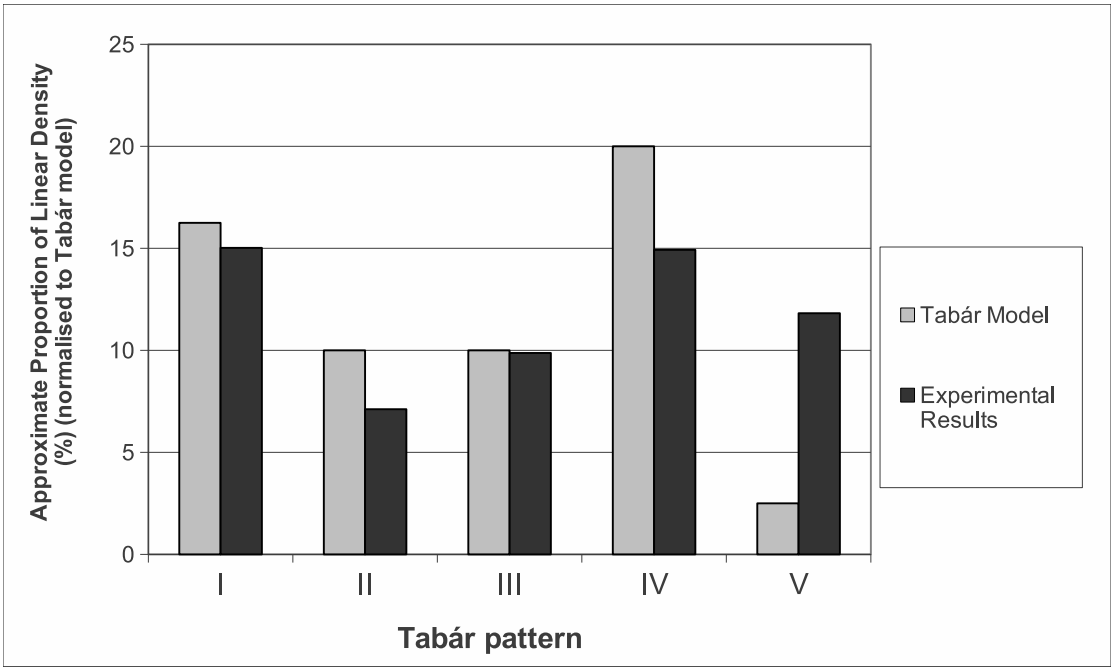


Figure 3.13: Graph showing the observed results at a threshold of 6/204 normalised by the mean of each data set (see Fig. 3.12).

3.3.3 Discussion and Conclusions

Whilst the proposed approach is simplistic, the results are promising and the analysis by Boyd SCC class demonstrates a clear ability to automatically distinguish between lower risk mammograms (classes 1–3) and higher risk mammograms (classes 4–6).

The observed proportions of linear structures correlate closely with Tabár's expected values for patterns I-IV, however the observed result for pattern V was considerably greater than expected (see Fig. 3.13). The analysis by Tabár pattern is interesting since the results for pattern V were unexpected. Tabár suggests that the linear density in pattern V mammograms should be very low [22] (see Fig. 3.12), whereas our results indicate a relatively high linear density for this pattern (see Figs. 3.9 and 3.13). However, a possible explanation might be that the line operator enhances linear structures even in dense tissue (see Figs. 3.6–3.8) and as such might result in a high proportion of linear structures, whereas under the Tabár classification this area might be assigned to one of the alternative

mammographic building blocks, such as fibrous tissue. Further work to segment and mask the fibrous tissue prior to application of the line detector is discussed in Sec. 3.4.

In addition, the results for Tabár pattern I mammograms, which demonstrate a high linear density, do not correlate with the results of low risk Boyd or BIRADS classes, which show low linear densities. Other studies have found that Tabár's patterns do not correlate well with other risk assessment models [29]. The principal anomaly is the low risk Tabár pattern I. The mammograms in our test set belonging to this pattern do not easily correlate with a particular Boyd classification, instead being spread amongst Boyd classes 2-5, with the majority seeming to belong to the high risk classes 4-5 [29].

The results of the analysis by BIRADS class are also less promising than the analysis by Boyd class, since it is not currently possible to distinguish between the low risk class 2 and the high risk class 4. We see in the analysis by Boyd class that the highest class (6) shows a slightly lower linear density than the classes immediately below it. At a threshold of 6/204 this decline is only slight and there is no significant difference between the linear densities of classes 4–6 (see Table 3.5). The density of class 6 is also significantly higher than all three of the low risk classes. The decline is accentuated when a threshold of 4/204 is taken, where we see a significant difference between classes 4 and 6. This indicates a dependence on the threshold value and a more principled approach to determine the threshold would be desirable.

Similarly, the highest risk BIRADS class (4) is indistinguishable at either threshold from the low risk class 2 based on this method, but is significantly lower than class 3. The trend seen in the results by Boyd class is also present in the BIRADS results. Owing to a lack of distinction between BIRADS classes 2 and 4, we are unable to make a reliable estimate on the BIRADS scale, and further

information will be necessary to achieve this. It is possible that a distinction may be achievable by applying the proposed method at a wider variety of scales, or alternatively additional data may be necessary, such as information relating to the distribution of the linear structures.

In summary, the proposed approach is promising but simplistic in that it considers only the density of linear structures and does not take in to account information relating to their distribution. The potential to automatically detect linear structures within a mammogram and relate these measurements to mammographic risk classifications has been demonstrated.

3.3.4 Investigating the Effect of Abnormalities

3.3.4.1 Background

During the investigation in section 3.3.1, concerns arose as to the effect of abnormalities in the data set. Since the MIAS database contains both normal and abnormal images, could the presence of abnormalities be affecting the results? The expected result was that the presence of abnormalities would not significantly affect the density of linear structures in the breast, since with the exception of spiculated lesions, the development of an abnormality is unlikely to increase or decrease the quantity of linear structures. It is quite likely, however, that the development of an abnormality would affect the distribution of these structures.

3.3.4.2 Method

Of the 321 MIAS images (see Sec. 3.2), 206 are normal (non-cancerous), and 115 are abnormal (cancerous).

The experiment described in section 3.3.1 was repeated on the subset of 206

normal images and the results compared with the results on the set of all images in order to determine whether the inclusion of abnormal images significantly affects the results. Graphs were produced showing the comparison of the two sets by both Tabár pattern and Boyd class (see Figs. 3.14 and 3.15).

3.3.4.3 Results and Discussion

The results of this analysis by Tabár pattern and Boyd SCC class are show in Figs. 3.14 and 3.15, respectively. The results show that there is no significant difference between the above threshold linearity in set of normal images compared with the complete dataset.

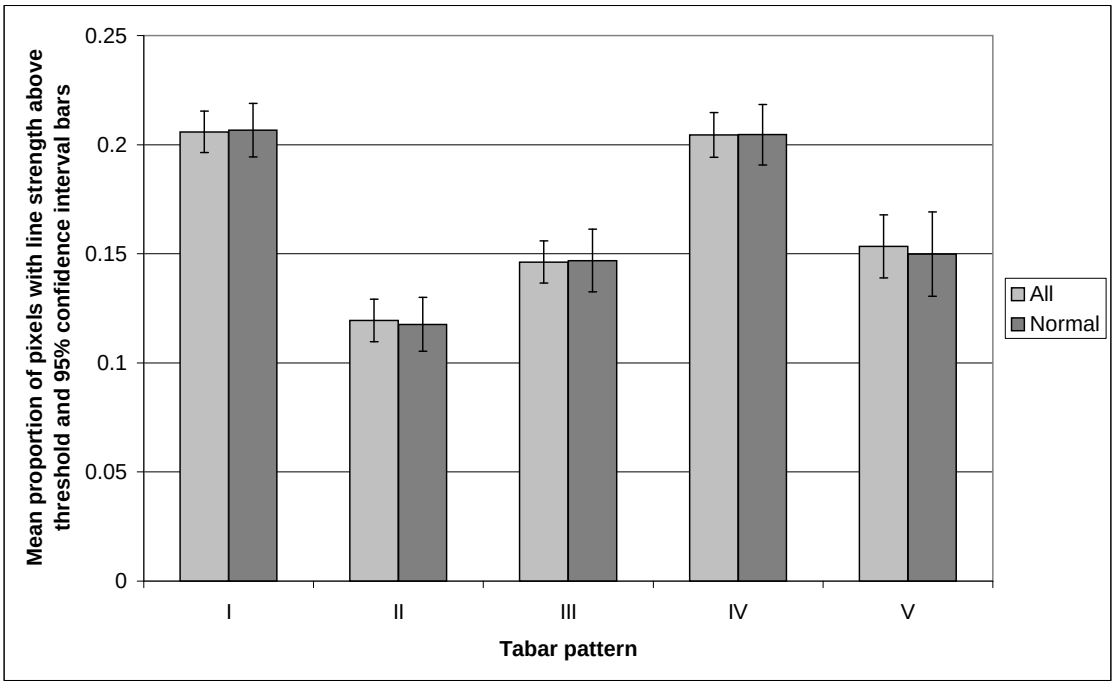


Figure 3.14: Graph showing the mean above threshold linearity and 95% confidence intervals of the mean for all images and the subset of normal images at each **Tabár** pattern.

The results of a comparison between the whole data set and the subset of normal images shows no significant differences between the two sets. This was the expected result (see Sec. 3.3.4.1). It is possible, however, that the presence of abnormalities may affect the distribution of the linear structures.

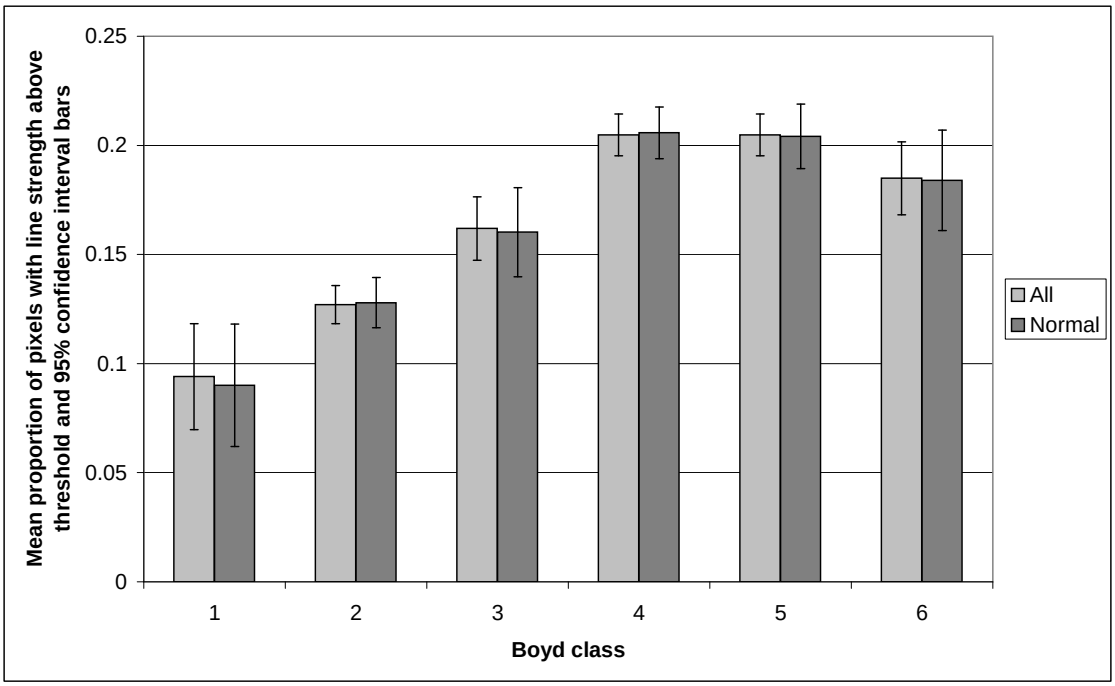


Figure 3.15: Graph showing the mean above threshold linearity and 95% confidence intervals of the mean for all images and the subset of normal images at each **Boyd** class.

3.4 Density Estimation

3.4.1 Background

Due to the anomaly observed regarding Tabár pattern V (see sections 3.3.2 and 3.3.3), a follow-up piece of work was undertaken which involved segmenting the dense from non-dense (fatty) tissue prior to running the line operator. The density of linear structures in each tissue type can then be analysed separately and should reduce the anomalous effects seen in the results for Tabár pattern V in the previous experiment.

The segmentation was conducted using a method described by Sivaramakrishna, *et al.* [81], which is based on Kittler and Illingworth's *Minimum Error Thresholding* [82]. The method involves converting the image to a *variation image*, which aims to differentiate between large bright areas (dense tissue) and small bright areas (e.g. ducts) prior to performing Kittler and Illingworth's opti-

mal threshold algorithm. The intention is to segment out large bright areas whilst leaving small bright areas.

3.4.2 Method

The method used was based on that used in section 3.3, and the experiment used the same 321 images from the MIAS database.

The method involves creating a variation image from the source image, which is then used as input to the Minimum Error Thresholding algorithm to produce an approximation of the dense tissue area. If the Minimum Error Thresholding algorithm were to be performed over the original image, the algorithm would place all bright areas of the image above the threshold. This would include dense tissue, but also smaller bright structures such as ducts and vessels. The purpose of the conversion to a variation image is to suppress small bright structures whilst enhancing large bright areas.

The variation image was calculated by dividing each pixel value by its corresponding unit variation V, given by

$$V(I,J) = \frac{1}{S}\sqrt{\left[\sum_{i=-N}^{N}\sum_{j=-N}^{N}\left(\frac{M(i,j) - M(I,J)}{M(I,J)}\right)^2\right]}, \tag{3.2}$$

over a square window with sides of length $(2N+1)$, where S is the number of pixels in the window. A window size of 5 (i.e. $N = 2$) was used for our experiment as suggested by the authors [81].

High pixel values in the original image that are associated with small bright areas will have a higher V value, and thus a lower value in the variation image than similar pixel values associated with large bright areas.

The variation image is then used as the source image for Kittler and Illingworth's Minimum Error Threshold algorithm [82]. This algorithm involves calculating a criterion function for each grey level value and finding its minimum. This is a relatively straightforward computation. The criterion function J for a possible grey level threshold value T is given by

$$\begin{aligned} J(T) = 1 \quad &+ \quad 2\left[P_1(T)\log\sigma_1(T) + P_2(T)\log\sigma_2(T)\right] \\ &- \quad 2\left[P_1(T)\log P_1(T) + P_2(T)\log P_2(T)\right]. \end{aligned} \tag{3.3}$$

The parameters $P_i(T)$, $\sigma_i(T)$ in (3.3), for grey level histogram $h(g)$ are given by

$$P_i(T) = \sum_{g=a}^{b} h(g) \tag{3.4}$$

and

$$\sigma_i(T) = \sqrt{\frac{\sum_{g=a}^{b}\left[g - \mu_i(T)\right]^2 h(g)}{P_i(T)}}, \tag{3.5}$$

where

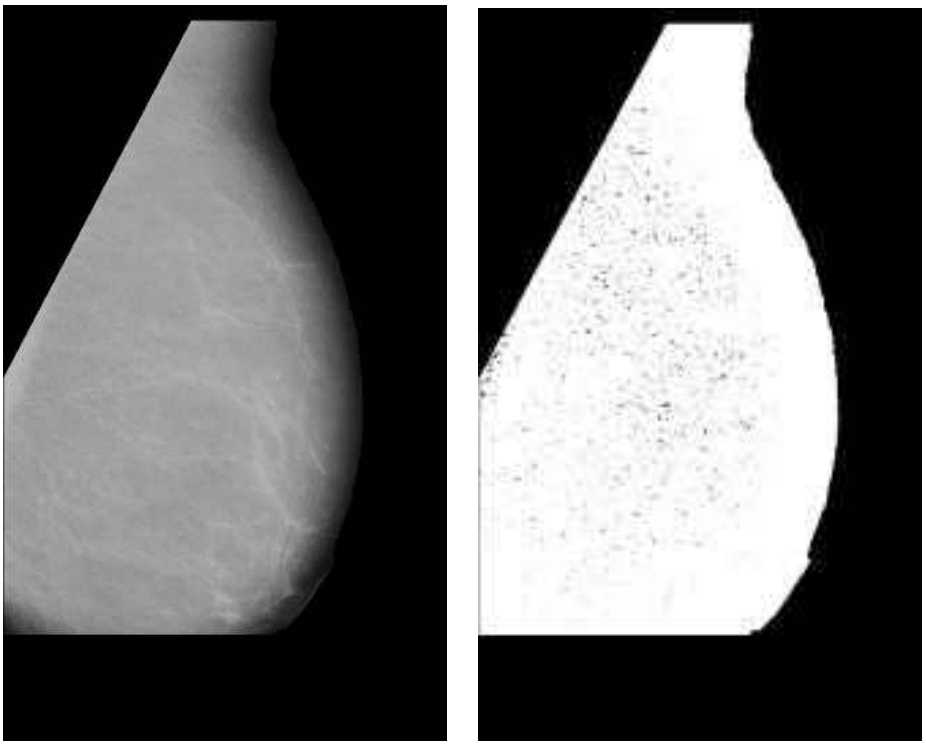
$$\mu_i(T) = \frac{\sum_{g=a}^{b} h(g)g}{P_i(T)}, \tag{3.6}$$

$$a = \begin{cases} 0 & i = 1 \\ T+1 & i = 2 \end{cases} \tag{3.7}$$

and

$$b = \begin{cases} T & i = 1 \\ n+1 & i = 2 \end{cases} . \tag{3.8}$$

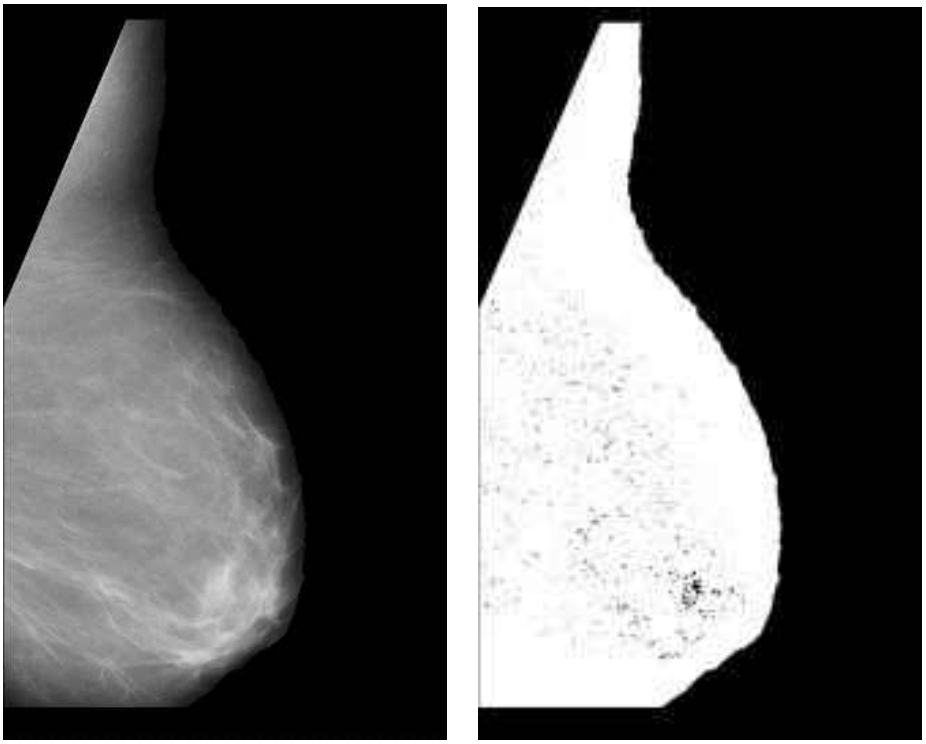
The minimum J was found and its corresponding value was used as the threshold for segmenting the variation image. An image mask was produced in order to easily identify the dense tissue areas in subsequent calculations. Some examples of the images and their resultant dense tissue masks are shown in Figs. 3.16–3.18.



(a) (b)

Figure 3.16: A typical mammogram of Boyd SCC class 1/Tabár pattern II/BIRADS class 1 (low risk). (a) – the original mammogram. (b) – the dense tissue mask. Areas segmented as dense tissue are shown in black (within the breast area), whilst areas segmented as non–dense (fatty) tissue are shown in white.

In addition to the segmentation of dense tissue, the original images were processed using Dixon and Taylor's Line Operator algorithm as used previously [63, 62] (see Sect. 3.3.1). The results were combined with the dense tissue mask in order to distinguish between lines in dense tissue and lines in non–dense (fatty)



(a) (b)

Figure 3.17: A typical mammogram of Boyd SCC class 3/Tabár pattern III/BIRADS class 2 (moderate risk). (a) – the original mammogram. (b) – the dense tissue mask. Areas segmented as dense tissue are shown in black (within the breast area), whilst areas segmented as non–dense (fatty) tissue are shown in white.

tissue.

The experiment resulted in a number of variables for each image:

(L_d) the number of pixels with above-threshold linearity in dense tissue,

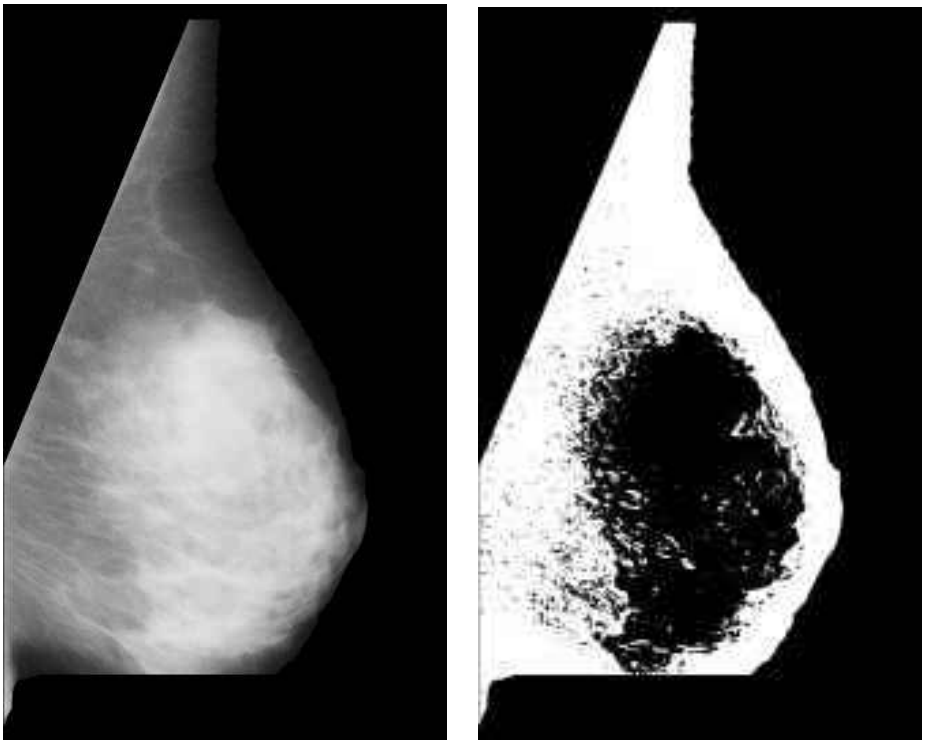
(L_f) the number of pixels with above-threshold linearity in fatty tissue,

(A_d) the total number of pixels in dense tissue, and

(A_f) the total number of pixels in fatty tissue.

This provides many combinations of interest, including

$$\mathbf{(a)}\frac{L_d}{A_d}, \mathbf{(b)}\frac{L_f}{A_f}, \mathbf{(c)}\frac{L_d}{A_d + A_f}, and \mathbf{(d)}\frac{L_f}{A_d + A_f}. \tag{3.9}$$



(a) (b)

Figure 3.18: A typical mammogram of Boyd SCC class 6/Tabár pattern IV/BIRADS class 3 (high risk). (a) – the original mammogram. (b) – the dense tissue mask. Areas segmented as dense tissue are shown in black (within the breast area), whilst areas segmented as non–dense (fatty) tissue are shown in white.

Following initial analysis, whereby the classification experiment was performed on a small subset of images, it was found that the results of Eq. 3.9 (a) and (b) yielded worse results than (c) and (d), so it was decided to focus on the latter two combinations. There are many combinations of the measured variables that could have been tried, however to investigate them all would have been infeasible and since (c) and (d) appeared to produce reasonable classification results, these were selected. The first two combinations consider only the density of linear structures in dense (or fatty) tissue as a proportion of the area of dense (or fatty) tissue, whereas the final two combinations consider the density of linear structures in dense (or fatty) tissue as a proportion of the total breast area. This is also affected by the relative proportions of dense and fatty tissue in the breast.

A number of problems were found with the density segmentation method. Some images produced criterion function graphs that were not bimodal, having

either zero or multiple minima. In addition, spurious values at the edges of the criterion function curves (caused by, for example, an image having no pixels with a greyscale value of 255) leading to false minima were initially a problem. This was overcome by selecting the lowest minimum where multiple minima occurred, and as a result of observations, by classifying the whole area as non-dense tissue where no minima occurred.

The segmentation was also found to be heavily dependent on the area used for segmentation, since if too much background area was included, in many non-dense breasts the threshold was placed between the background tissue and the breast tissue, resulting in the whole of the breast area being classified as dense tissue. To overcome this, a threshold was set and all pixels with greyscale values below this threshold were ignored. This effectively removed any surrounding background tissue from the analysis.

3.4.3 Results

An overview of the differences between Tabár patterns, Boyd classes and BIRADS classes is shown in graph form in Figs 3.19, 3.20 and 3.21, respectively. In order to present a more robust analysis of significant differences between patterns and classes, Mann–Whitney U tests were performed on each pair of patterns and classes, and tables of these results are shown in Tabs 3.7, 3.8 and 3.9, respectively.

Results are shown by Tabár pattern (see Fig. 3.19 and Tab. 3.7), Boyd class (see Fig. 3.20 and Tab. 3.8) and BIRADS class (see Fig. 3.21 and Table 3.9). The results by Tabár pattern and BIRADS class show that there are significant differences between all patterns or classes, both in dense and fatty tissue, although the differences are more significant in dense tissue than in fatty tissue. The results by Boyd class show significant differences between all classes in dense tissue and between some classes in fatty tissue, however some Boyd class combinations in

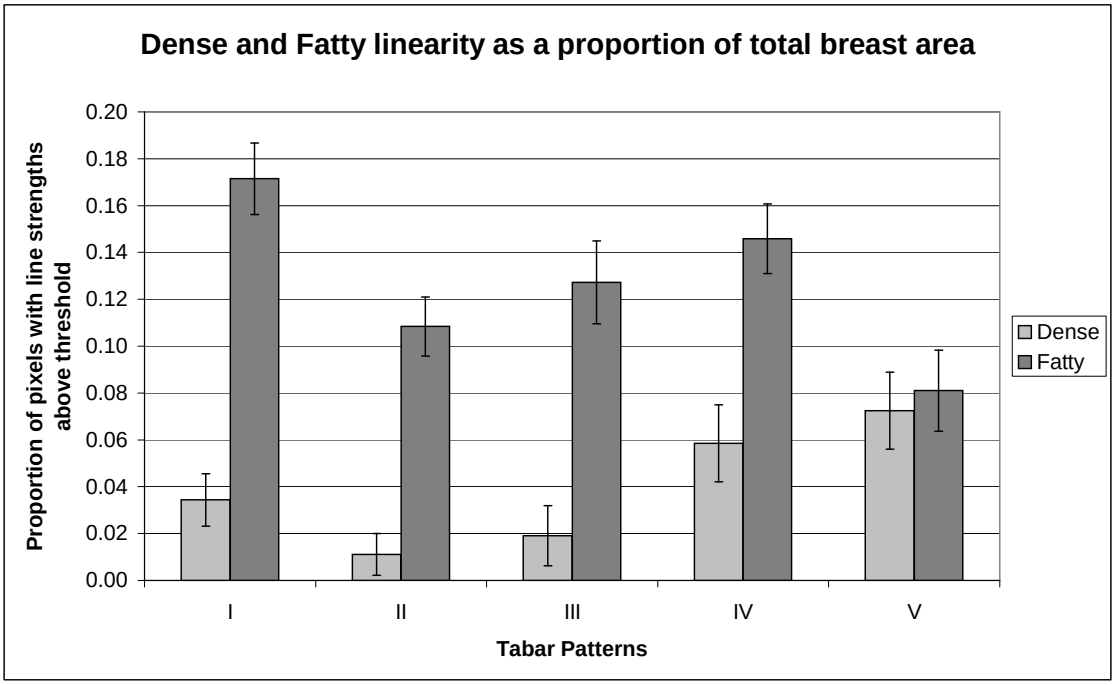


Figure 3.19: Graph showing the mean proportion of pixels with line strengths above a threshold $T = 4/204$ and 95% confidence intervals of the mean for images of each **Tabár** class. Results are shown for linear structures in dense and fatty tissue as a proportion of total breast area.

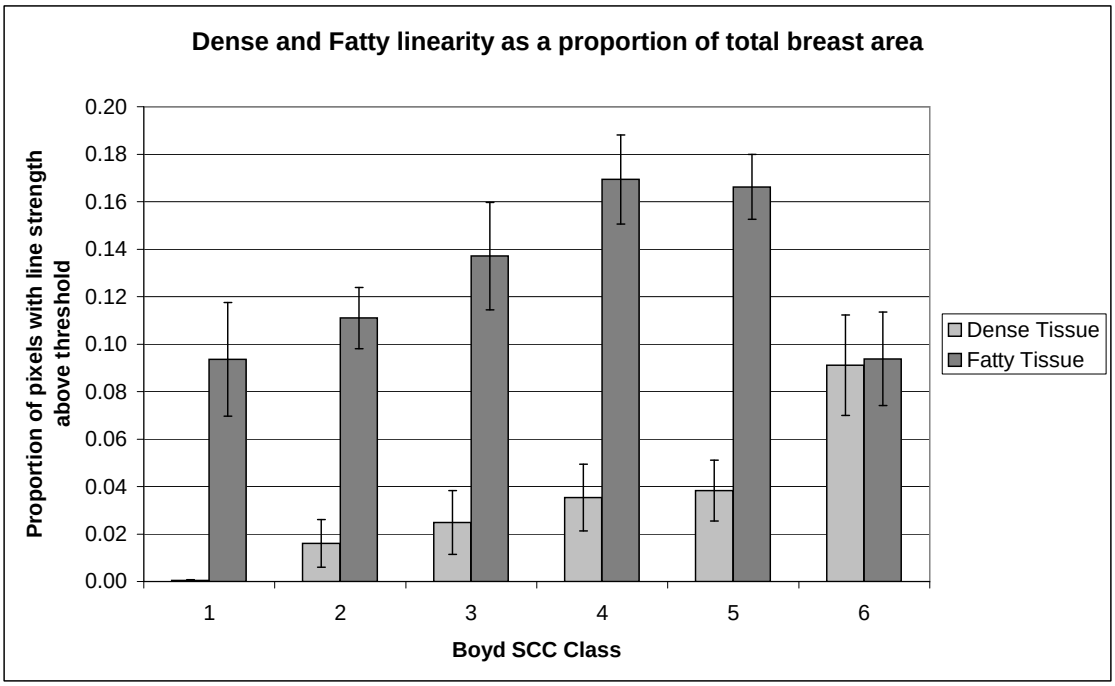


Figure 3.20: Graph showing the mean proportion of pixels with line strengths above a threshold $T = 4/204$ and 95% confidence intervals of the mean for images of each **Boyd** class. Results are shown for linear structures in dense and fatty tissue as a proportion of total breast area.

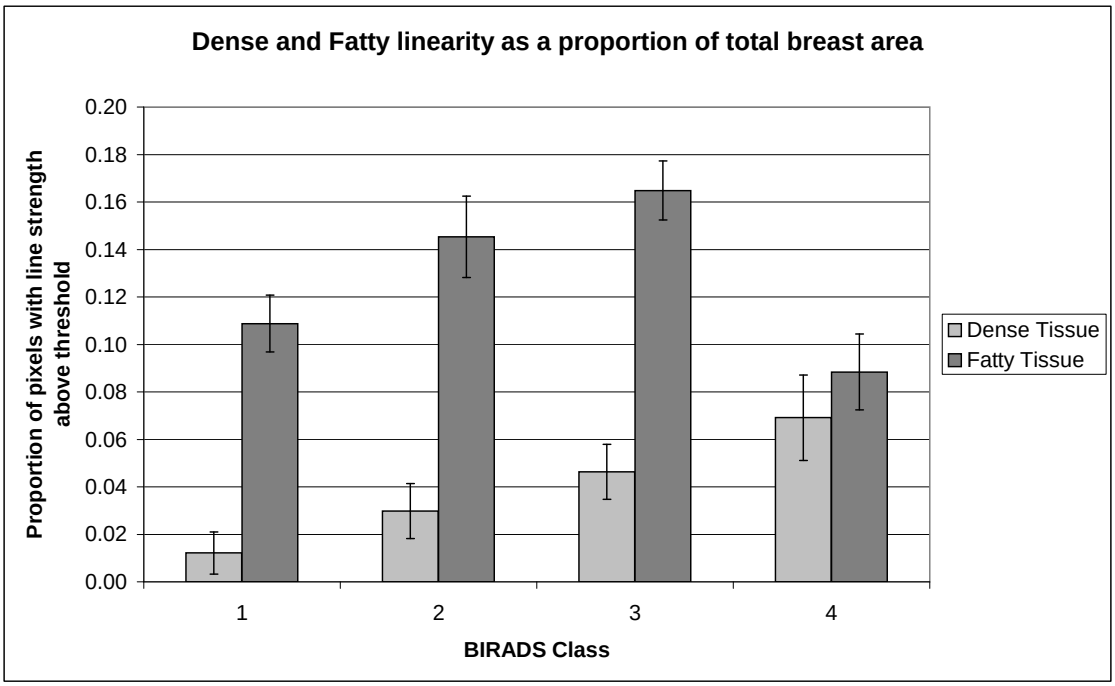


Figure 3.21: Graph showing the mean proportion of pixels with line strengths above a threshold $T = 4/204$ and 95% confidence intervals of the mean for images of each **BIRADS** class. Results are shown for linear structures in dense and fatty tissue as a proportion of total breast area.

Table 3.7: The p-values obtained by Mann-Whitney tests on each combination of **Tabár** patterns for linear structures in dense and fatty tissue, at a threshold $T = 4/204$. All results are significant at $\alpha = 0.05$.

Tabár	**II**	**III**	**IV**	**V**
I	0.0000	0.0000	0.0000	0.0000
II		0.0003	0.0000	0.0000
III			0.0000	0.0000
IV				0.0027

Linear structures in dense tissue.

Tabár	**II**	**III**	**IV**	**V**
I	0.0000	0.0000	0.0032	0.0000
II		0.0101	0.0001	0.0089
III			0.0297	0.0010
IV				0.0000

Linear structures in non–dense (fatty) tissue.

Table 3.8: The p-values obtained by Mann-Whitney tests on each combination of **Boyd SCC** classes for linear structures in dense and fatty tissue, at a threshold $T = 4/204$. Results not significant at $\alpha = 0.05$ are shaded.

Boyd	**2**	**3**	**4**	**5**	**6**
1	0.0204	0.0006	0.0001	0.0000	0.0001
2		0.0000	0.0000	0.0000	0.0000
3			0.0005	0.0000	0.0000
4				0.0005	0.0000
5					0.0000

Linear structures in dense tissue.

Boyd	**2**	**3**	**4**	**5**	**6**
1	0.1997	0.0734	0.0161	0.0036	0.5018
2		0.0101	0.0000	0.0000	0.0585
3			0.0024	0.0382	0.0104
4				0.1458	0.0000
5					0.0000

Linear structures in non–dense (fatty) tissue.

Table 3.9: The p-values obtained by Mann-Whitney tests on each combination of **BIRADS** classes for linear structures in dense and fatty tissue, at a threshold $T = 4/204$. All results are significant at $\alpha = 0.05$.

BIRADS	**2**	**3**	**4**
1	0.0000	0.0000	0.0000
2		0.0000	0.0000
3			0.0000

Linear structures in dense tissue.

BIRADS	**2**	**3**	**4**
1	0.0001	0.0000	0.0448
2		0.0410	0.0001
3			0.0000

Linear structures in non–dense (fatty) tissue.

fatty tissue do not show significant differences.

A threshold of 4/204 was used for the analysis during this experiment, since higher thresholds resulted in no or very few pixels in dense tissue being above the threshold for many images, making calculations and statistical analysis difficult, whilst lower thresholds were found to include substantial levels of background noise.

3.4.3.1 A Comparison of The Observed Results in Fatty Tissue with Tabár's Expected Results

Figure 3.22 shows the observed results for area of linear structures in non–dense (fatty) tissue normalised to Tabár's expected results (see Fig. 3.12) at pattern II. The result for pattern V has improved considerably from the result from the initial experiment (see Fig. 3.13).

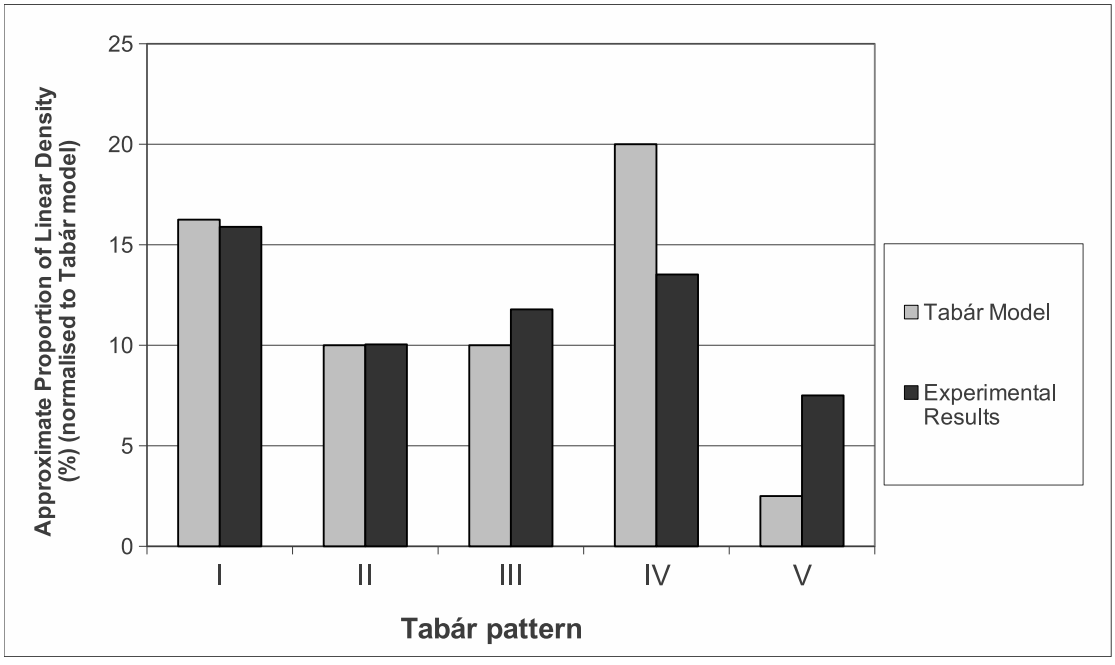


Figure 3.22: Graph showing the observed results in fatty tissue normalised by pattern II to Tabár's expected results (see Fig. 3.12).

3.4.4 Discussion and Conclusions

The proposed approach demonstrates the ability to distinguish between all Tabár patterns and Boyd and BIRADS classes using the area of linear structures in dense tissue as a proportion of total breast area (see Tabs 3.7, 3.8 and 3.9). The result by Boyd class was as expected, since it is well known that the proportion of dense tissue is related to mammographic risk [14, 15, 16, 17, 18, 19]. Since the Boyd metric is an indirect measure of breast density it can be assumed from this result that as the proportion of dense tissue increases, the area of linear structures in dense tissue also increases.

The results by Tabár pattern and BIRADS class also demonstrate a similar pattern. As with the results by Boyd class, we also see a positive relationship between linearity in dense tissue and risk when compared to BIRADS classes and Tabár patterns II-V (see Figs. 3.19, 3.20 and 3.21). The results for the area of linear structures in fatty tissue shows a slightly different pattern, where there is a general positive relationship between linearity and risk in low and moderate risk categories, followed by a sharp reduction in linearity in the highest risk categories (see Figs. 3.19, 3.20 and 3.21). This effect can be explained by the prevalence of dense tissue in the highest risk mammograms, meaning that there is little fatty tissue and as a result, fewer linear structures in fatty tissue.

The overall results for linear density in non–dense (fatty) tissue do not demonstrate such potential as for those in dense tissue. Whilst it is possible to distinguish between all Tabár patterns and BIRADS classes using the area of linear structures in fatty tissue, the ability to distinguish between Boyd classes using this measurement is more limited (see Tabs 3.7, 3.8 and 3.9). Where it is possible to distinguish between patterns and classes, the *p–values* generally indicate that the differences are less significant using linear structures in fatty tissue as a measurement rather than linear structures in dense tissue.

A disadvantage of the dense tissue segmentation method used is that it may be subject to variations between images. The dense tissue thresholding is carried out on each image individually, meaning that it is possible that variations in brightness, contrast and other image properties will affect the outcome of the thresholding. Mammograms taken on different machines and at different times, and processed differently are likely to be of slightly different brightness and contrast. This is a difficult problem to overcome. A single global threshold would eliminate differences caused by different proportions of tissue types in different images. However, if a single threshold were to be chosen for a set of images being processed, this would cause undesirable effects when processing images of different brightness and contrast caused by different machines and other variables.

A remaining issue in the linear structure detection algorithm is that the line strength threshold is currently chosen experimentally, and in this experiment a threshold of 6/204 was found to result in unusable results for many images. A more principled approach to determining the threshold would be an improvement, and future work will investigate the possibility of using the grey scale histogram to automatically determine a suitable threshold.

The result for Tabár pattern V when the linear structures in dense tissue are removed is closer to Tabár's expected value than the result obtained during the initial experiment prior to tissue segmentation (see Figs 3.22 and 3.12, respectively). This lends credibility to the hypothesis that the initial result was too high because linear structures were being detected in dense tissue whereas in Tabár's model this entire area would be assigned to a different building block. However, although the results obtained from the new experiment now show pattern V as having the lowest proportion of linear density, the result remains higher than Tabár's expected value. In addition, the segmentation has caused the proportion of linear density in pattern IV to be reduced below Tabár's expected value.

In summary, the results are highly promising and the proposed method provides a possible approach for the automatic classification of mammograms in to Tabár patterns, Boyd or BIRADS classes. The ability to automatically classify mammograms by a common risk assessment metric would be an extremely useful aid to radiologists that may lead to improved early detection rates. It is intended for future work to investigate the automatic risk classification of mammograms.

3.5 Classification

3.5.1 Background

Having demonstrated that the density of linear structures in the breast can be automatically detected and correlated with the expected proportions for risk patterns (see Secs 3.3–3.4), an experiment was proposed to investigate whether mammograms can be automatically placed in to risk categories using this information. The main purpose of this work is to investigate whether the use of linear structure information in an automatic risk classifier leads to improved accuracy.

Initial experiments using a simple linear classifier were conducted and then developed using a more complex non–linear support vector machine–based classifier.

3.5.2 Risk Classification using Linear Techniques

3.5.2.1 Method

Three hundred and twenty-one mammographic images from the Mammographic Image Analysis Society (MIAS) database [23] were classified according to BIRADS classes [27] by three expert radiologists (see section 3.2). Example images of low,

moderate and high risk mammograms are shown in Fig. 3.23 (a).

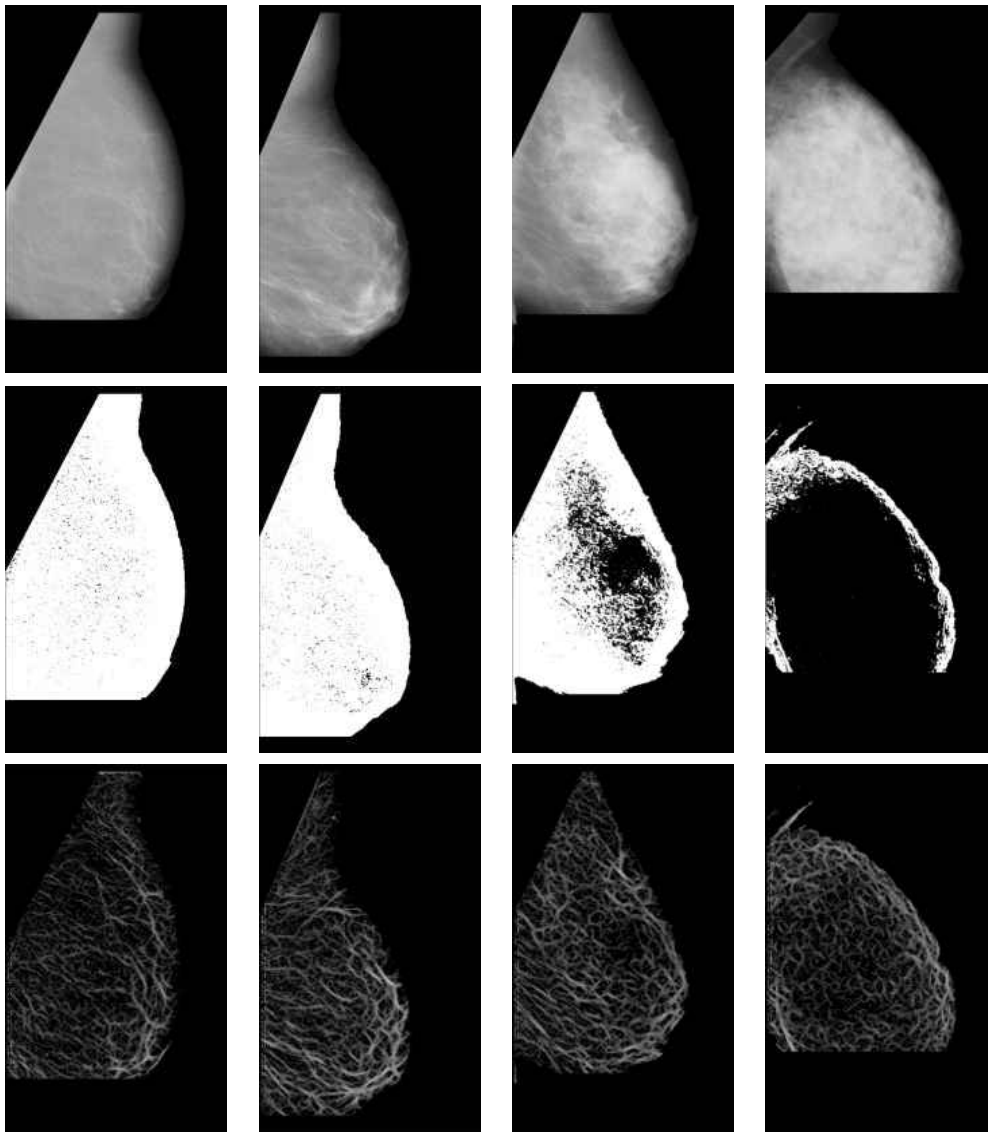


Figure 3.23: Some typical mammograms of various BIRADS classes. The left column shows a mammogram of BIRADS class 1 (low risk), the second column shows a mammogram of BIRADS class 2 (low–moderate risk) and the third column shows a mammogram of BIRADS class 3 (moderate–high risk), and the final column shows a mammogram of BIRADS class 4 (high risk). The images in row (a) show the original mammograms, row (b) shows the results of density segmentation (non-dense tissue is shown in white), and row (c) shows the results after processing with the line operator. The images have had a mask applied to remove everything outside the breast area. The lines in (c) have been enhanced for viewing.

All 321 images were segmented using the method described in Sec. 3.4. This produced masks identifying the dense tissue in the image and allowed for the consideration of linear structures in each tissue type independently. Examples of

density masks produced are shown in Fig. 3.23 (b). The results of this segmentation produced two values used in later calculations:
– the total number of pixels in dense tissue: A_d,
– the total number of pixels in fatty tissue: A_f.

In addition, the images were processed using the line operator method described in Sec. 3.3), producing a measurement of line strength at each pixel. Fig. 3.23 (c) shows examples of low, moderate and high risk mammograms following processing with the line operator. When considered alongside the density mask two further values were produced for use in further calculations:
– the number of pixels with above-threshold linearity in dense tissue: L_d,
– the number of pixels with above-threshold linearity in fatty tissue: L_f.

Finally, the images were classified in BIRADS classes using linear discriminant analysis (see Sec. 3.5.2.1.1) based on the features extracted from the processed images and the results analysed.

3.5.2.1.1 Classification Classification was conducted using linear discriminant analysis, which takes several factors as predictors and attempts to classify each mammogram in to BIRADS classes. Classification was conducted three times – firstly using density information only, secondly using linear structure information only, and a third time using both density and linear structure information.

Linear discriminant analysis uses one or more predictors, which are used to attempt to distinguish between objects of the various classes. In the classification based on density information, the following values were used as predictors:

$$\frac{A_d}{A_d + A_f}, \log \frac{A_d}{A_d + A_f}, \tag{3.10}$$

where $\frac{A_d}{A_d+A_f}$ represents the area of dense tissue (A_d) as a proportion of the total

breast area ($A_d + A_f$).

During experimentation it was found that adding the log of the value as an additional predictor increased the performance of the classifier. It can be assumed that the log function accentuates the lower parts of the scale where most of the mammograms lie, producing greater separation between classes. The classification based on linear structure information used the following values as predictors:

$$\frac{L_d}{A_d}, \frac{L_f}{A_f}, \log \frac{L_d}{A_d}, \log \frac{L_f}{A_f}, \tag{3.11}$$

where $\frac{L_d}{A_d}$ represents the linear density in dense tissue and $\frac{L_f}{A_f}$ represents the linear density in non–dense tissue. Figs. 3.24–3.25 show plots of these values grouped into BIRADS classes.

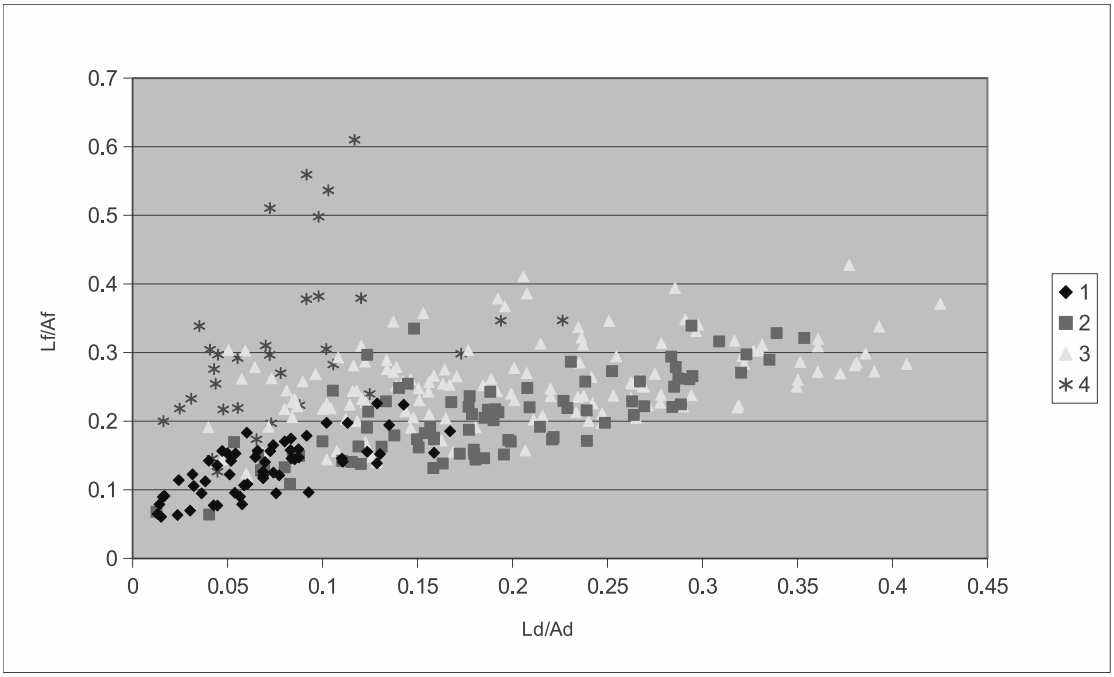


Figure 3.24: Graph showing a plot of $\frac{L_d}{A_d}$ against $\frac{L_f}{A_f}$ grouped into BIRADS classes.

The classification based on both linear structure and density information uses a single classifier which incorporates all of the above six values as predictors (i.e. the predictors from both the density classification and linear structure classification are used in the third classification).

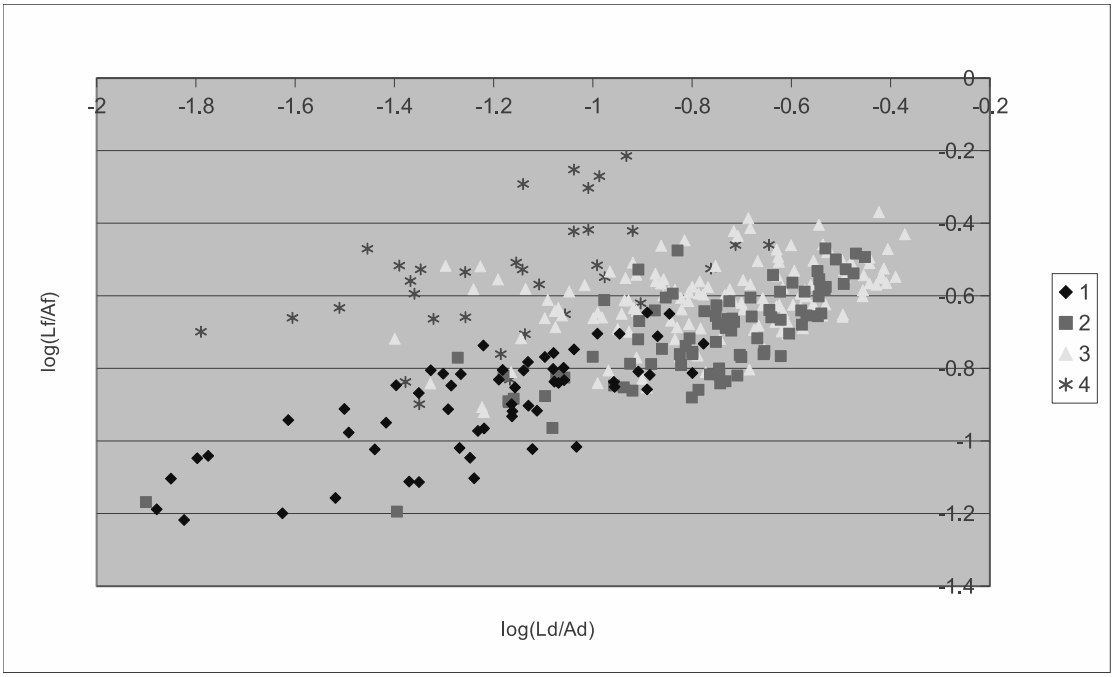


Figure 3.25: Graph showing a plot of $log\frac{L_d}{A_d}$ against $log\frac{L_f}{A_f}$ grouped into BIRADS classes.

Results of each classification are shown in table form, including the proportion of correct classifications and the kappa (κ) coefficient [99, 100]. The κ coefficient is a means of estimating agreement in categorical data, and is given by

$$\kappa = \frac{P(D) - P(E)}{1 - P(E)} \tag{3.12}$$

where $P(D)$ is the proportion of correct classifications and $P(E)$ is the proportion expected by chance. The κ coefficient is a means of estimating agreement in categorical data and can be used to assess the performance of the classification. A list of common interpretations of κ values is shown in Tab. 3.10 [100].

Table 3.10: Common interpretations of κ values [100].

κ	Agreement
< 0	Poor
$0 - 0.20$	Slight
$0.21 - 0.40$	Fair
$0.41 - 0.60$	Moderate
$0.61 - 0.80$	Substantial
$0.81 - 1.00$	Almost Perfect

3.5.2.2 Results

3.5.2.2.1 Classification Based on Density Results of the leave–one–woman–out linear discriminant analysis by BIRADS class based on density information are shown in Tab. 3.11. The results indicate a 66% accuracy in classifying the mammograms. The κ coefficient indicates moderate agreement between the classified results and the true classes.

Table 3.11: Classification results by BIRADS class based on density information.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	**62**	14	2	0
2	18	**57**	19	4
3	4	28	**68**	9
4	3	4	5	**24**
Proportion correct	0.713	0.553	0.728	0.649
	0.657			
κ **coefficient**	0.54			

3.5.2.2.2 Classification Based on Linear Structures Results of linear discriminant analysis by BIRADS class based on linear structure information are shown in Tab. 3.12. The results indicate a 67% accuracy in classifying the mammograms. The κ coefficient indicates moderate agreement between the classified results and the true classes.

Table 3.12: Classification results by BIRADS class based on linear structure information.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	**68**	10	7	1
2	9	**71**	27	0
3	9	19	**50**	9
4	1	3	10	**27**
Proportion correct	0.782	0.689	0.532	0.730
	0.673			
κ **coefficient**	0.56			

3.5.2.2.3 Classification Based on Linear Structures and Density Results of linear discriminant analysis by BIRADS class based on a combination of density and linear structure information are shown in Tab. 3.13. The results indicate a 72% accuracy in classifying the mammograms. The κ coefficient indicates a substantial agreement between the true classes and the placed classes. These results show that a combined approach is superior to using either of the single components.

Table 3.13: Classification results by BIRADS class based on density and linear structure information.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	**70**	12	2	1
2	12	**73**	25	0
3	2	15	**60**	9
4	3	3	7	**27**
Proportion correct	0.805	0.709	0.638	0.730
	0.717			
κ **coefficient**	0.62			

3.5.2.3 Discussion and Conclusions

The results are promising and show a substantial improvement between classification using density information and classification using combined information, leading to the conclusion that linear structure information is valuable in the automatic risk classification of mammograms.

An alternative approach to the analysis might be to look at whether the mammograms are correctly classified in to high/low risk groups. Results of this analysis for combined density and linear structure information are shown in Tab. 3.14. It can be seen that the classifier provides 84% correct high/low classification ($\kappa = 0.79$) based on both density and linear structure information.

A comparison with other methods described in literature shows that the pro-

Table 3.14: Classification results by BIRADS class based on density and linear structure information summarised as low (classes 1-2)/high (classes 3-4) groups.

Placed Class	**True Class**	
	Low	**High**
Low	**167**	28
High	23	**103**
Proportion correct	0.879	0.786
	0.841	
κ **coefficient**	0.79	

posed method performs well, achieving 72% correct classification. Oliver *et al.* [101] performed texture–based classification by BIRADS class on the MIAS images used in this analysis. This classification achieved 66% accuracy when classifying the images without prior segmentation. Results were improved significantly to between 79% and 82% correct classification when the images were segmented in to regions prior to the texture analysis as first suggested by Karssemeijer [102].

The alternative classifications described in literature suggest that texture–based classification with prior segmentation performs better than density–based segmentation [97]. As such, an option for future work would be to investigate classification based on this approach combined with linear structure information. The combination technique used was a straightforward one, however optimisation techniques exist which may provide improved results for the combined classification.

For the purposes of a comparison between density, linear structure and combined information the methods used were adequate, however more accurate methods for the segmentation of dense tissue and classification have been demonstrated [101, 97]. The results shown demonstrate that the inclusion of linear structure information improves automatic risk classification when compared to a conventional density–based classifier.

3.5.2.4 Application of Non–Linear Classification Techniques

The simple classification experiments were followed–up with non–linear classification using a Support Vector Machine (SVM) based classifier [103]. The classifier used a radial basis function (RBF) kernel, and a grid search of parameters was performed to obtain suitable parameters for each experiment.

The experiment was performed on two independent datasets, one of which has been divided in to two groups. Initial results were obtained using the Mammographic Image Analysis Society (MIAS) database [23] (see section 3.2). A second dataset containing 831 mammograms from the Digital Database for Screening Mammography (DDSM) database [104] was used to confirm results. This dataset consisted of digitised mammograms from 831 different patients classified according to the BIRADS metric. Figure 3.26 shows some typical examples from the DDSM database, alongside corresponding dense tissue masks and linear structure detection results. The images used had been reduced in size and padded to fill a 1024x1024 matrix. Table 3.15 shows the distribution of images according to the BIRADS classification. In order to avoid bias in the results when comparing the two datasets, the left images from the MIAS dataset were discounted, resulting in a dataset consisting of 160 images from individual women.

The threshold used for the MIAS database remained 6/204, however since the DDSM database images differed in size and contrast the threshold selection procedure described in Sec. 3.3.1 was repeated to find a suitable threshold for these images. This resulted in a threshold of 2/204 being selected as the DDSM images were found to contain less contrast than the MIAS images, which resulted in lower line strength measures. Future work may involve the development of a more principled approach to determining the threshold using the grey–level histogram of a dataset.

Six classification experiments were performed, corresponding with the LDA experiments. Cross–validation was used in order to provide unbiased results. This was performed in a leave–one–out methodology on the MIAS dataset, and owing to computational constraints, in a 10–fold methodology on the larger DDSM dataset.

BIRADS class	**1**	**2**	**3**	**4**
Images	106	336	255	134

Table 3.15: The distribution of DDSM database images according to the BIRADS classification metric.

3.5.2.4.1 Results Results using the non–linear Support Vector Machine based classifier experiment are shown as confusion matrices in Tabs 3.16–3.21. Tables 3.16 and 3.17 provide results for classification based on density information, Tabs 3.18 and 3.19 provide results for classification based on linear structure information, and Tabs 3.20 and 3.21 provide results for classification based on both density and linear structure information. Table 3.22 shows the SVM classification results grouped by low/high risk categories. Results for the MIAS database have been summed across all leave–one–woman–out classifications and results for the DDSM database have been summed across the ten folds.

Table 3.16: Classification based on density information using the MIAS dataset.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	35	11	3	2
2	4	13	10	1
3	2	24	28	7
4	2	5	4	9
Proportion correct	0.8140	0.2453	0.6222	0.4737
	0.5312			
κ **coefficient**	0.37			

3.5.2.4.2 Discussion and Conclusions Whilst the use of simple LDA analysis produces results that compare less favourably with some other studies [93, 91, 97], a pattern is evident throughout both datasets. The classification results suggest that the density–based classifier produced the least accurate overall results,

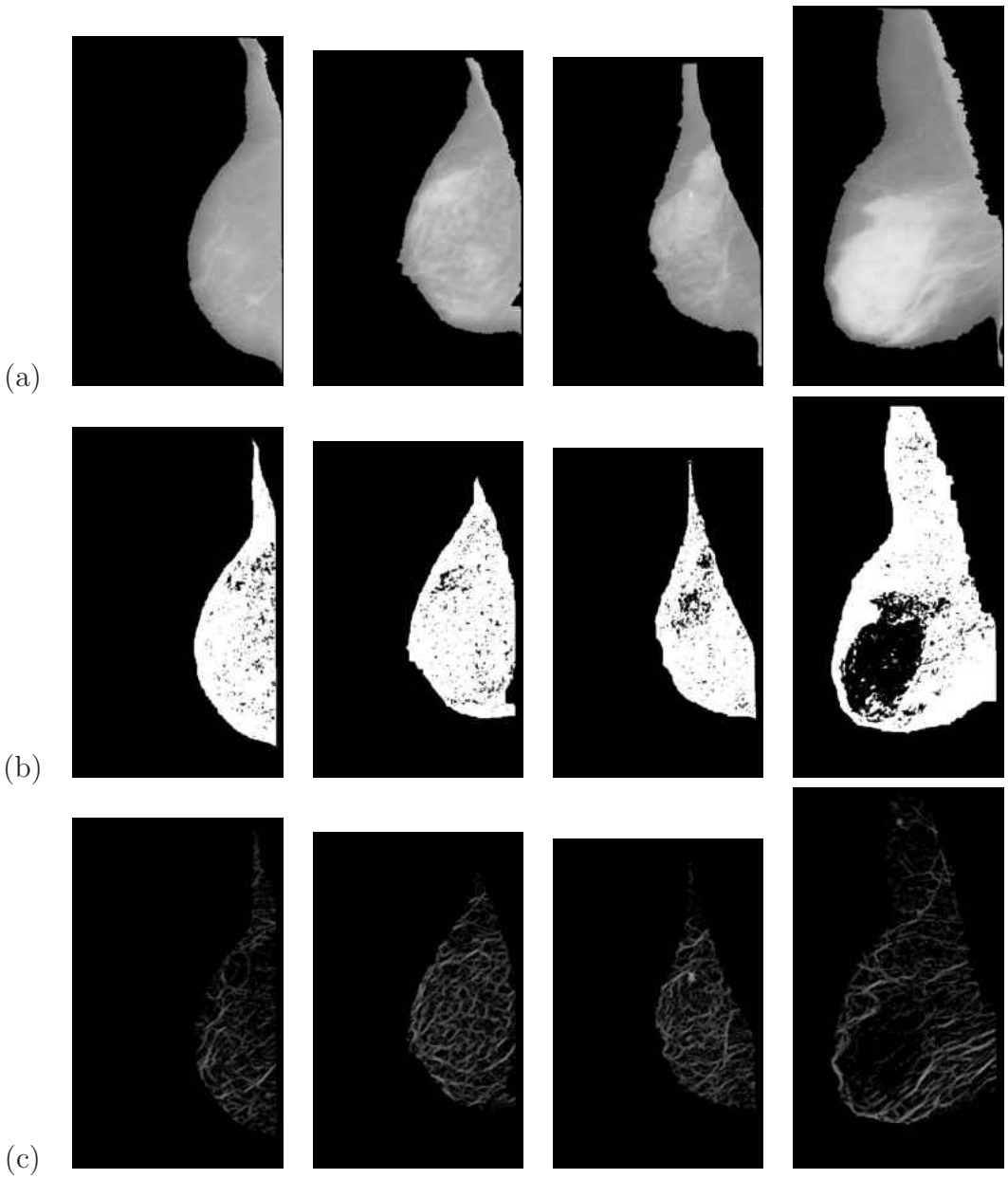


Figure 3.26: Some typical mammograms from the DDSM database of each BI-RADS class. The left column shows a mammogram of BIRADS class 1 (low risk), the second column shows a mammogram of BIRADS class 2 (low–moderate risk) and the third column shows a mammogram of BIRADS class 3 (moderate–high risk), and the final column shows a mammogram of BIRADS class 4 (high risk). The images in row (a) show the original mammograms, row (b) shows the results of density segmentation (non-dense tissue is shown in white), and row (c) shows the results after processing with the line operator. The images have had a mask applied to remove everything outside the breast area. The lines in (c) have been enhanced for viewing.

Table 3.17: Classification based on density information using the DDSM dataset.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	12	7	0	0
2	91	278	90	34
3	3	49	143	79
4	0	2	22	21
Proportion correct	0.1132	0.8274	0.5608	0.1567
	0.5463			
κ **coefficient**	0.40			

Table 3.18: Classification based on linear structure information using the MIAS dataset.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	29	27	3	1
2	9	34	14	0
3	4	12	28	6
4	1	0	0	12
Proportion correct	0.6744	0.6415	0.6222	0.6316
	0.6438			
κ **coefficient**	0.53			

Table 3.19: Classification based on linear structure information using the DDSM dataset.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	15	8	0	0
2	88	264	74	28
3	2	60	158	69
4	1	4	23	37
Proportion correct	0.1415	0.7857	0.6196	0.2761
	0.5704			
κ **coefficient**	0.43			

Table 3.20: Classification based on density and linear structure information using the MIAS dataset.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	33	5	1	2
2	8	40	12	0
3	1	8	31	3
4	1	0	1	14
Proportion correct	0.7674	0.7547	0.6889	0.7368
	0.7375			
κ **coefficient**	0.65			

Table 3.21: Classification based on density and linear structure information using the DDSM dataset.

Placed Class	**True Class**			
	1	**2**	**3**	**4**
1	14	5	0	0
2	90	270	76	32
3	2	57	161	69
4	0	4	18	33
Proportion correct	0.1321	0.8036	0.6314	0.2463
	0.5752			
κ **coefficient**	0.43			

Table 3.22: SVM classification results divided in to high/low risk groups.

Classification	**Dataset**	**BIRADS Class**			κ coefficient
		Low	**High**	**All**	
Density	MIAS	0.6563	0.7500	0.6938	0.59
	DDSM	0.8778	0.6812	0.7858	0.71
Linear structures	MIAS	0.8229	0.7188	0.7813	0.71
	DDSM	0.8484	0.7378	0.7966	0.73
Density & linear	MIAS	0.8958	0.7656	0.8438	0.79
	DDSM	0.8575	0.7224	0.7942	0.73

the linear structure–based classifier performed slightly better, and the density and linear structure information produced the most accurate classification. This suggests that linear structure information is of value as part of a mammographic risk classifier. The SVM–based classifier produced results that were generally improvements over those produced by LDA, however the same trend is demonstrated in these results.

A notable trend in the results was the difference between the two datasets. The DDSM dataset produced considerably poorer classification results than the MIAS dataset. An important factor that may have contributed to this effect is the accuracy of the ground truth BIRADS classifications. Whereas the MIAS dataset has been classified by three expert radiologists and consensus classifications were found, the circumstances of how the DDSM dataset has been classified are unclear. It is possible that only a single radiologist was used, and it is also possible that there was no consistency of radiologist over the entire dataset. Since a large

degree of inter-radiologist variation is often observed[105], even amongst expert radiologists, it is possible that the "true" classifications for the DDSM dataset are not as accurate as for the MIAS dataset. Other studies have observed less accurate classification results using the DDSM dataset compared with the MIAS dataset [97].

In addition to the breakdown of results according to individual BIRADS class, a common usage by radiologists in practice is a straightforward high/low risk categorisation. Table 3.22 shows the SVM classification results as high/low groups, where BIRADS classes 1 and 2 are considered low risk and classes 3 and 4 are considered high risk.

An interesting outcome of this analysis is that the results for the two datasets are broadly comparable when the BIRADS classes are grouped together, suggesting that much of the poor performance of the DDSM classification may be due to more variable ground truth classifications. This analysis also demonstrates that a high degree of accuracy can be obtained using the proposed method when classifying mammograms in to high/low risk categories.

A final consideration is the extraction of data from images, where the proposed approach considers only the density of linear structures and does not take in to account information relating to their distribution. Since several risk assessment models are based on the parenchymal patterns in the breast [14, 22], it has been suggested that the distribution of linear structures may be related to risk, and as such may provide additional information useful for risk assessment. An investigation in to this is described in Chapter 4.

Chapter 4

3D Risk Assessment

4.1 Introduction

Conventional two-dimensional projection mammography plays a very significant role in breast cancer detection, diagnosis and treatment. However, it is well–known that 2D mammography has several inherent limitations, caused by the projection of the 3D breast anatomy on to a 2D plane. These include cancers being obscured by superimposed normal tissue and overlapping normal tissue creating the artificial appearance of densities [56]. These limitations often result in false–negative or false–positive diagnoses, increasing risk to the patient or exposing them to stress and unnecessary, often painful follow–up procedures.

Three–dimensional breast tomosynthesis provides a significant advance over projection mammography. Tomosynthesis effectively eliminates the effects of superimposed tissue on parenchymal structures of interest [25, 26]. This can increase margin visibility, especially in dense breasts and has been shown to improve lesion visibility [24] and the visibility of parenchymal texture features [60].

Experiments described in this chapter investigate the feasibility of extract-

ing information relating to the distribution of linear structures from 3D breast tomosynthesis images and using this data to provide automatic risk assessment.

4.2 Detection of Anatomical Linear Structures in Raw Tomosynthesis Images

4.2.1 Introduction

The purpose of this work is to investigate the sensitivity of the detection of linear structures in raw tomosynthesis images compared to conventional mammograms.

Breast tomosynthesis acquires a series of projection x–ray images as the x–ray source moves in an arc around the fixed breast and digital imaging detector. With the exception of their acquisition angle, the 'raw' projection images are similar to conventional x–ray mammograms. However they are taken using a significantly lower x–ray dose than that used for conventional mammograms, such that the overall dose received by the patient is similar for the two methods and the tomosynthesis patient is not exposed to increased levels of harmful x–ray radiation [24]. This has the effect of limiting the number of images acquired so as to maintain a reasonable level of contrast within each image, leading to potential difficulties in the three–dimensional reconstruction.

As a result of the lower x–ray dose, a principal concern was that there would be insufficient contrast available to make reliable detection of linear structures possible. In order to ascertain whether sufficient information was available in tomosynthesis images, an investigation was conducted using the linear structure detection methodology developed in Chapter 3 and applying it to both conventional mammograms and the central image from a sequence of raw tomosynthesis

images captured during the same session.

Throughout this experiment raw tomosynthesis images were used (the two–dimensional projection images) rather than the three–dimensional reconstructed images because they are more easily comparable with conventional mammograms due to their dimensionality. Additionally, the raw tomosynthesis images allowed for a preliminary investigation to be undertaken using existing detection methodology prior the development of a three–dimensional linear structure detector.

4.2.2 Data

Conventional mammograms and tomosynthesis images were captured for 40 patients during the same screening sessions. For each patient, two conventional mammograms (left and right breast) and two sets of nine raw tomosynthesis images (in an arc around the patient) were captured using a GE Senographe 2000D FFDM system at the University of Pennsylvania. The system was modified for the series of nine projection tomosynthesis images to be taken across a -25° – +25° range at 6.25° intervals. Projection images were acquired with a pixel resolution of 0.1mm.

Of the sets of nine raw tomosynthesis images, the central image represented a side–view and corresponded most closely with the MLO conventional mammograms, and as such these central images were selected for use in our experiment. The images were represented as 12–bit grey–scale, 1914x2294 pixel DICOM files.

Figures 4.1 and 4.2 show a conventional mammogram and the sequence of raw tomosynthesis images for the same breasts.

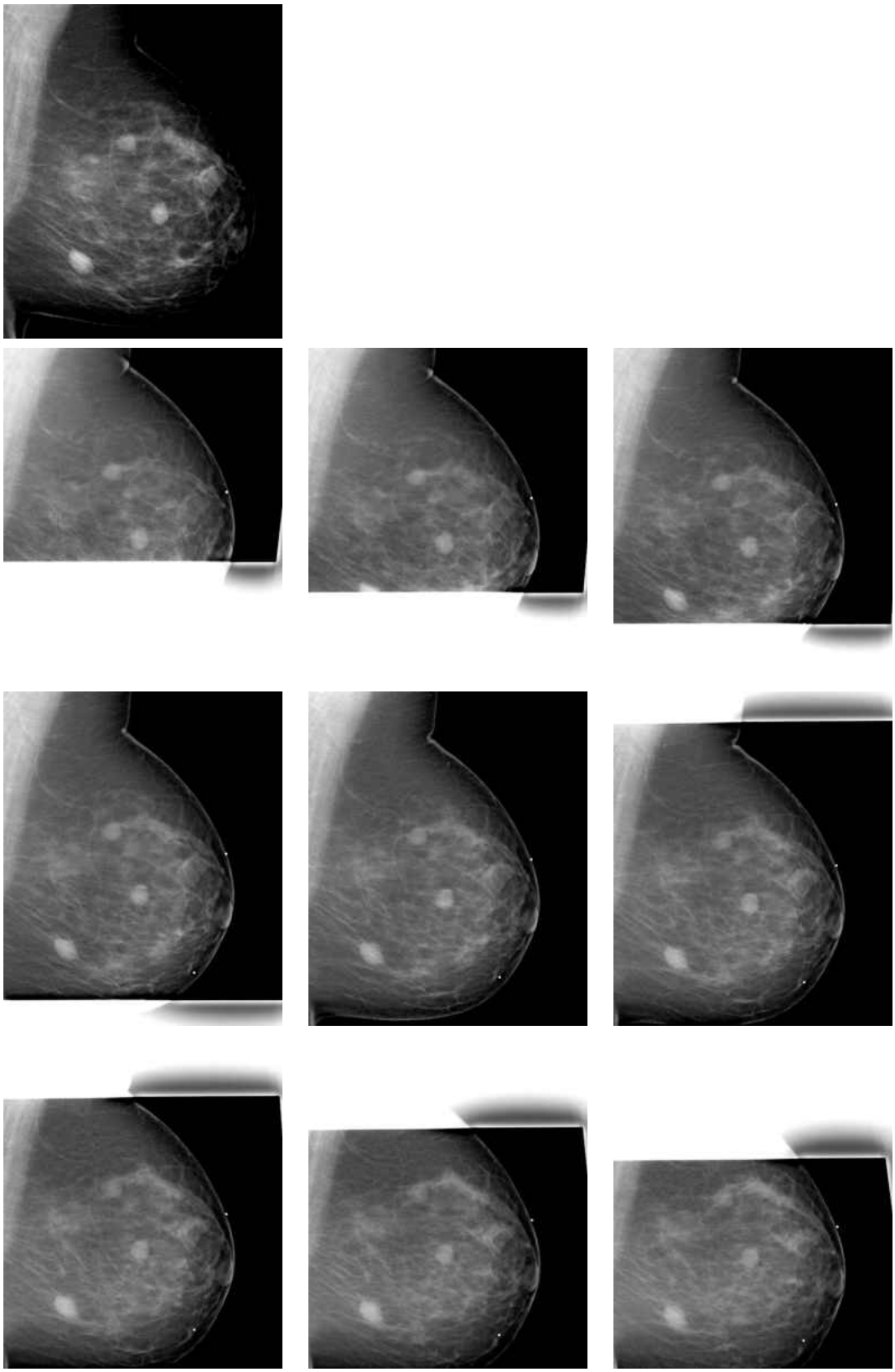


Figure 4.1: A conventional mammogram (top left) and a corresponding sequence of raw tomosynthesis images.

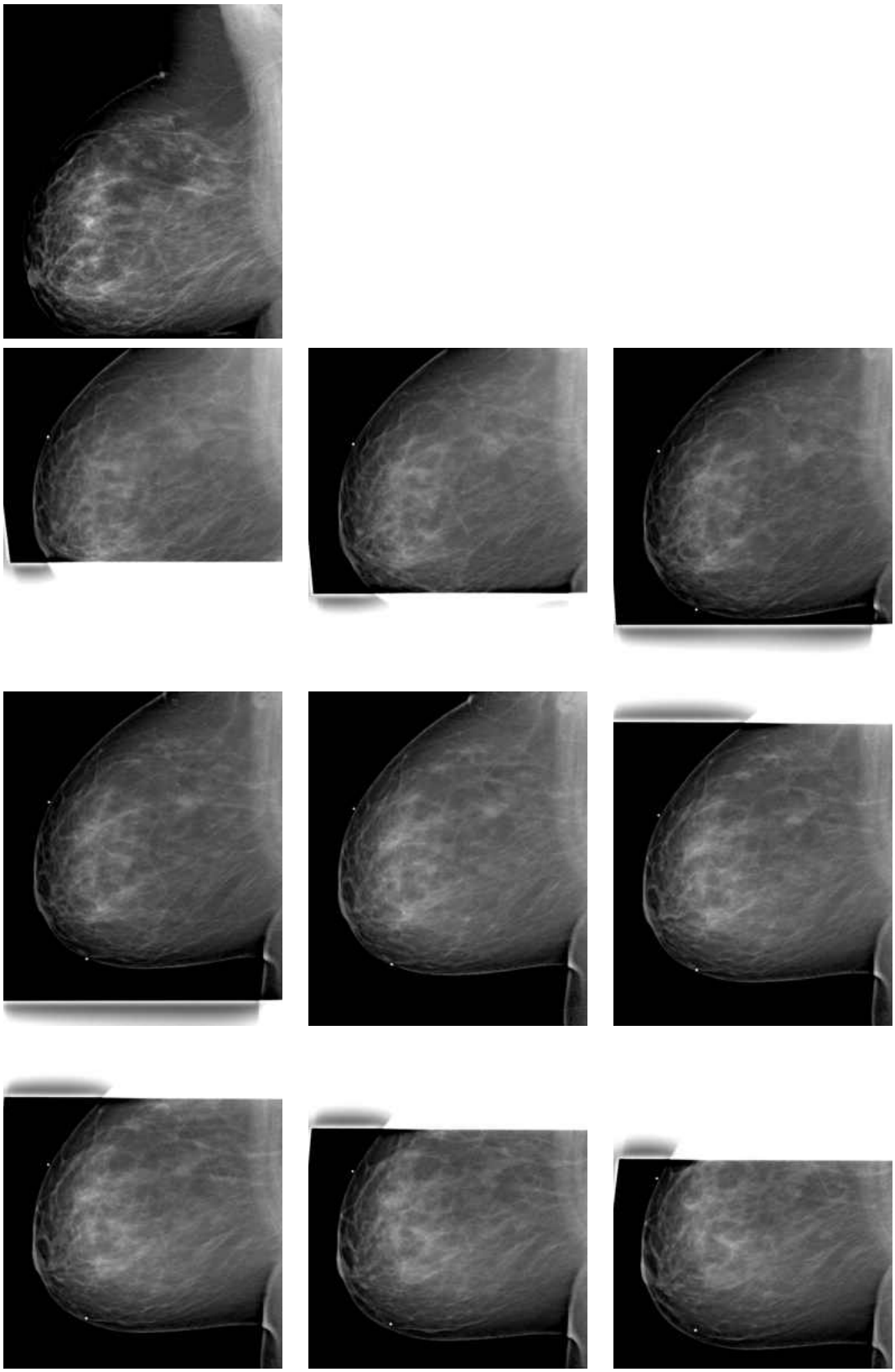


Figure 4.2: A conventional mammogram (top left) and a corresponding sequence of raw tomosynthesis images.

4.2.3 Method

The Line Operator method (see Sec. 3.3.1) was applied to each of the 80 mammograms and corresponding central raw tomosynthesis images. Figure 4.3 shows the original mammogram and raw tomosynthesis images from Fig. 4.1 and the results of processing with the line operator.

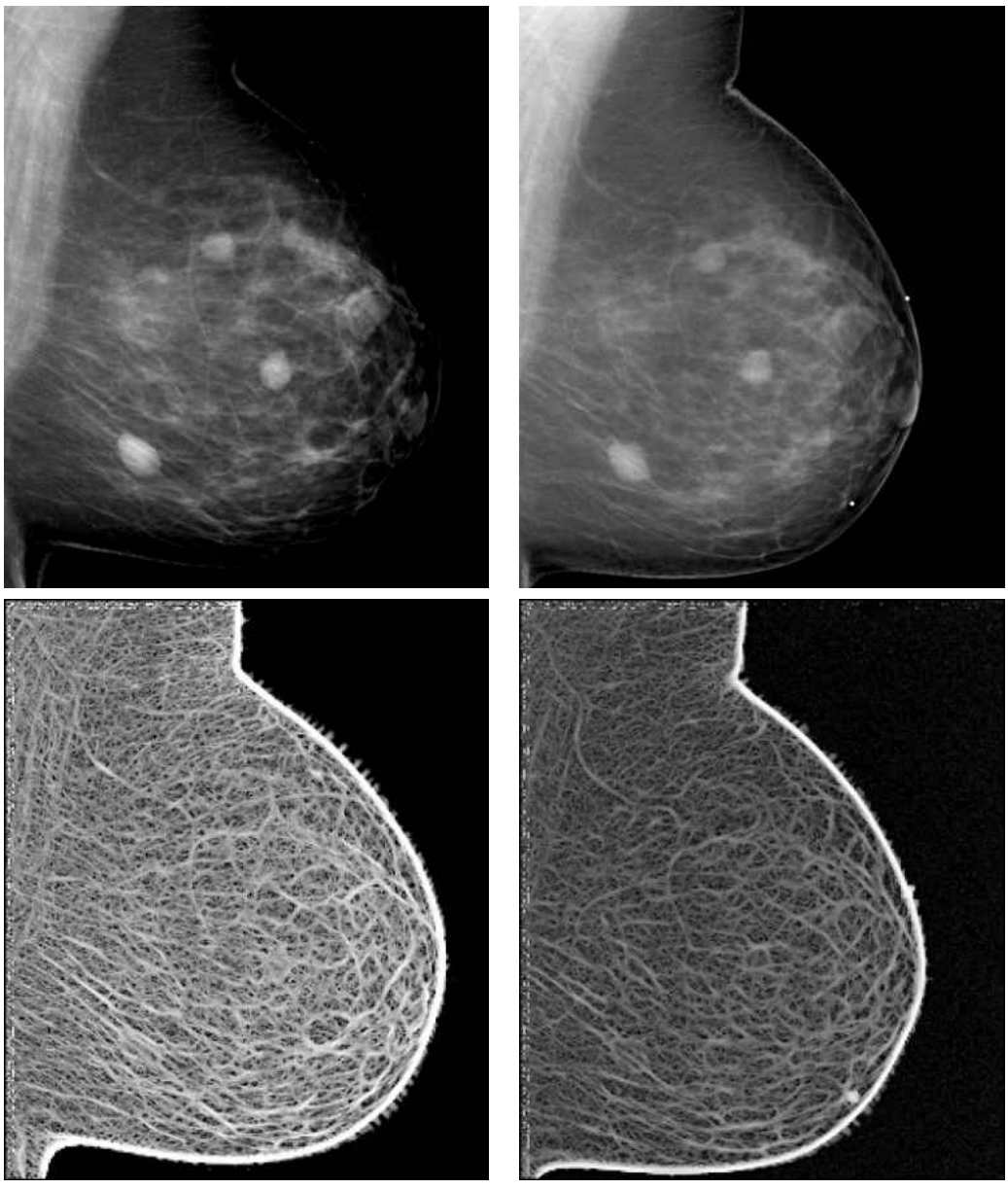


Figure 4.3: Line operator results showing a conventional mammogram (left) and a corresponding raw tomosynthesis image (right). The linear structures have been enhanced for display.

Following application of the line operator, the line strength results were thresh-

olded to remove background tissue, aiming to leave only linear structures. The choice of threshold greatly affects the measure of proportion of linear structures. Initial experiments involved performing the linear structure detection method on a small, randomly–selected subset of images of each set at a wide range of thresholds and selecting two thresholds that visually appeared to remove most background tissue whilst leaving most of the linear structure information intact. The two thresholds that were used during this experiment were 16/4096 and 20/4096. Whilst different thresholds could have been selected for each set of images, these thresholds appeared to be suitable for both the mammograms and central tomosynthesis images.

Line strength values in these images were in the range 0–4096 (the maximum grey–scale value for the images used). The two sets of images were thresholded at each threshold by removing all pixels with line strength values below that threshold. Figure 4.4 shows the conventional mammogram and raw tomosynthesis images shown in Fig. 4.3 following thresholding at 16/4096 and 20/4096.

4.2.4 Results

The graph in Fig. 4.5 shows the proportion of above–threshold linearity in raw tomosynthesis images against the corresponding conventional mammograms. A threshold of 20/4096 was used in both the conventional mammograms and tomosynthesis images in order to achieve the best correlation. This achieved a Pearson product moment correlation coefficient of 0.866.

Figure 4.6 shows the proportions of above threshold linearity in raw tomosynthesis images and mammograms at all combinations of thresholds used.

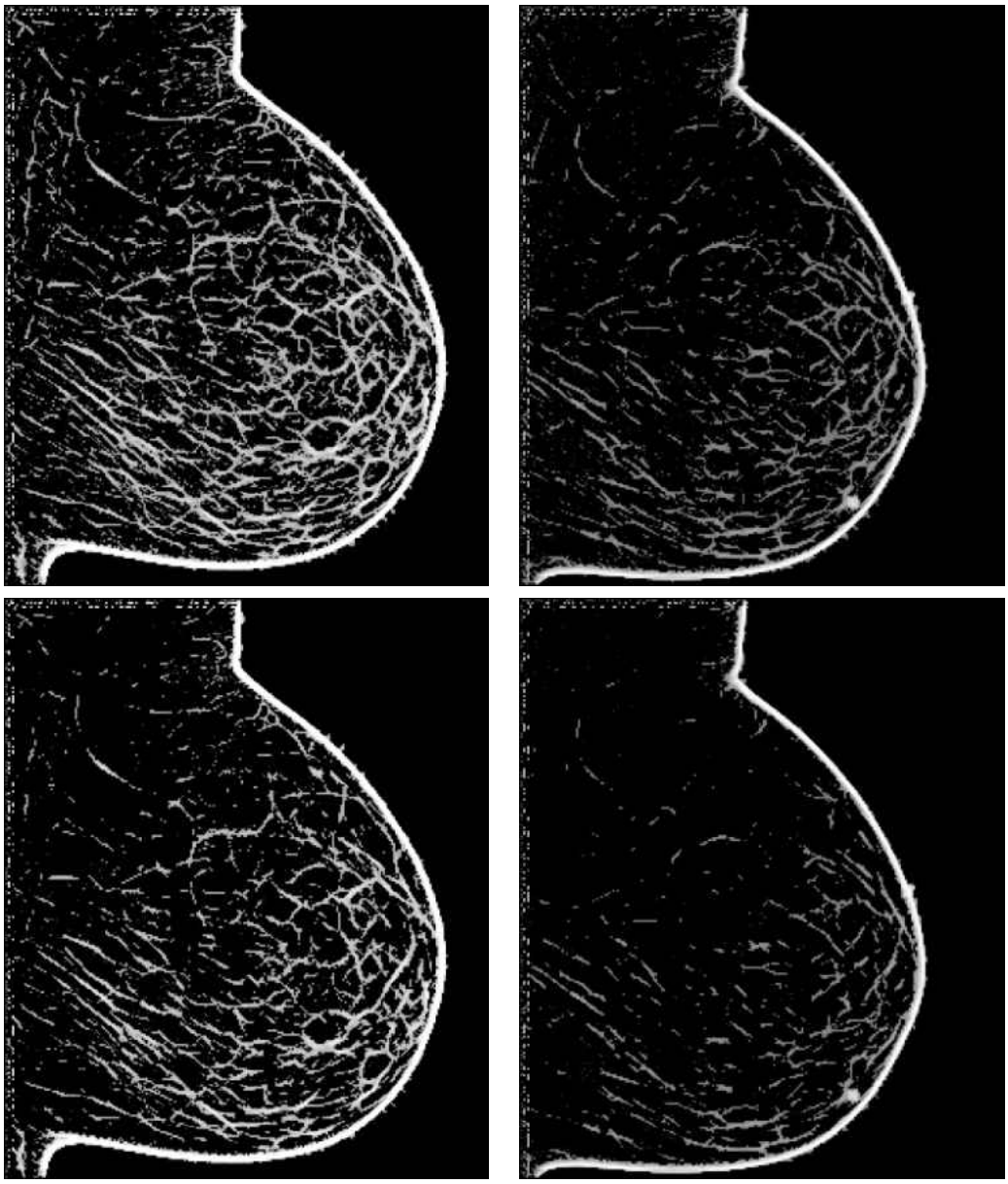


Figure 4.4: Thresholded line operator results showing a conventional mammogram (left) and a corresponding raw tomosynthesis image (right). The top images have been thresholded at 16/4096 and the bottom images have been thresholded at 20/4096. The linear structures have been enhanced for display.

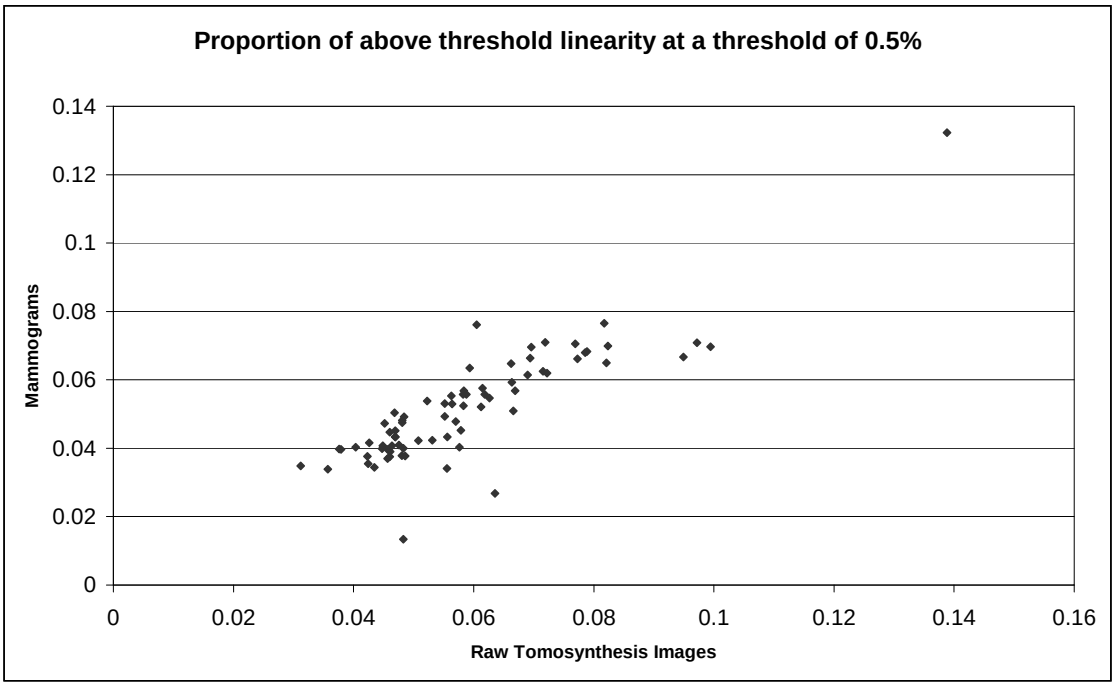


Figure 4.5: A graph showing the proportion of above–threshold linearity in raw tomosynthesis images against corresponding conventional mammograms at a threshold of 20/4096.

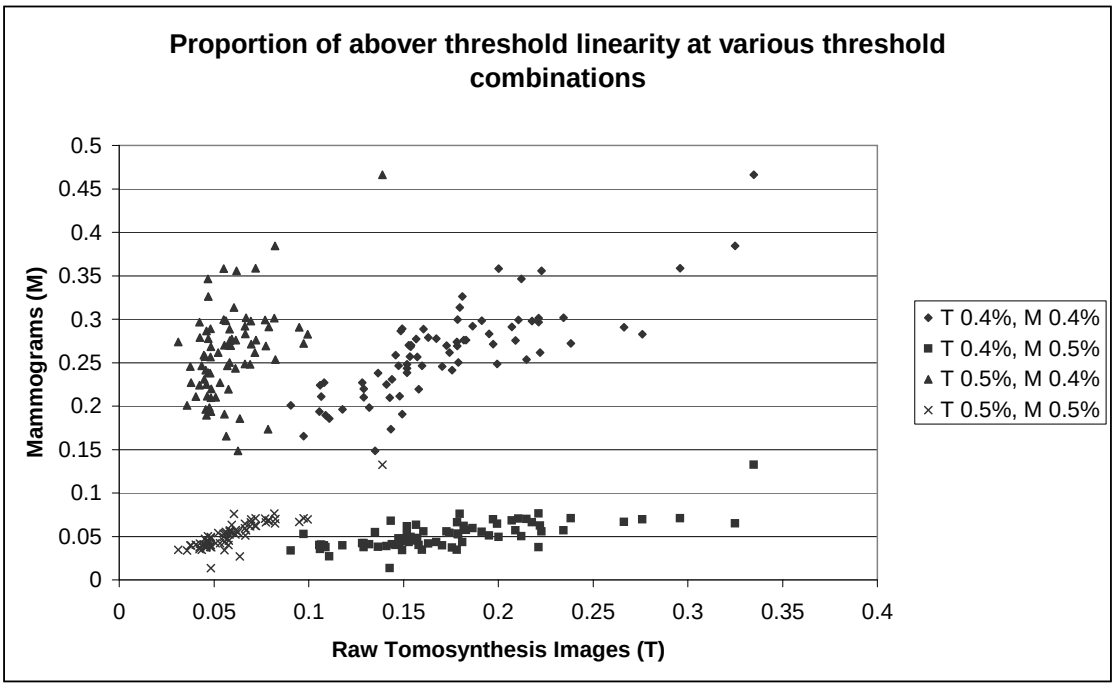


Figure 4.6: A graph showing the proportion of above–threshold linearity in raw tomosynthesis images against corresponding conventional mammograms at various combinations of thresholds. In the legend, T indicates the threshold used for the raw tomosynthesis images and M indicates the threshold used for mammograms.

4.2.5 Discussion and Conclusions

Linear structure information in mammograms has been linked to mammographic risk and detected linear structure information has shown promise in improving the specificity of automatic risk assessment [22, 106]. Tomosynthesis offers the potential for significant improvements over mammograms by eliminating the problems associated with the representation of the 3D anatomy in 2D [25, 26, 24]. The results shown in Fig. 4.5 show a high correlation between the detected proportions of linear structures in raw tomosynthesis images compared with conventional mammograms (0.866). This suggests that the linear structure information in tomosynthesis images may be useful in mammographic risk assessment investigations.

The selected threshold affected the correlations. Since the contrast varied between the different image types, the thresholds were selected independently in order to achieve the closest correlation, however the correlations were highest when both were thresholded at 20/4096 (0.866) and when both were thresholded at 16/4096 (0.813). A larger sample of thresholds would need to be used in order to assess whether this is significant. The correlation remained good when mammograms thresholded at 20/4096 were compared with tomosynthesis images thresholded at 16/4096 (0.714), however the correlation was less acute when these thresholds were reversed (0.475).

It is suggested that since the contrast was greater in the mammograms (due to the higher x–ray dose) the line strength values tended to be higher, and as a result when the mammogram results were thresholded at a low proportion less background would be removed. When a high threshold was applied, more background would be removed and a large amount of linear structure information remained. However, when the lower–contrast tomosynthesis image results were thresholded at a high proportion more linear structure information would be removed (see Fig. 4.4). When these combinations or thresholds are compared, the

correlation would be expected to be poorer than for other combinations.

Overall, the results are promising because they suggest that, whilst the correlation is not exact and the linear structures are clearly less easily detectable in the raw tomosynthesis images, the correlation across the dataset remains very high and as such it could be expected that the information would represent a useful indicator in risk assessment.

4.3 3D Linear Structure Detection

4.3.1 Background

It has been demonstrated that the linear structure information captured for tomosynthesis images is detectable and is likely to be useful in indicating mammographic risk (see Sec. 4.2), so the next step towards producing a classifier for tomosynthesis images is to demonstrate a line detector for the 3–dimensional reconstructed image volumes. The purpose of this work is to show that linear structures can be detected in tomosynthesis volumes and to investigate how the detected linear structure information correlates with linear structure information detected in 2–dimensional mammograms.

4.3.2 Data

The raw tomosynthesis images described in Sec. 4.2.2 were used to reconstruct 3–dimensional image volumes using a filtered–backprojection approach. The volumes were represented with slices at 1mm intervals and an in–plane pixel resolution of 0.22mm. Typical volumes consisted of 50 – 90 slices. Images from 39 women were used in this study due to corruption in the image data for one patient.

Figure 4.7 shows a typical mammogram and slices from the corresponding tomosynthesis volume.

4.3.3 Method

The 2–dimensional mammograms and 3–dimensional tomosynthesis volumes were processed using linear structure detection algorithms. The method used on mammograms (see Sec. 3.3.1) was adapted for use on 3–dimensional tomosynthesis volumes.

The original method calculates a measure of line strength and orientation for each pixel in a 2–dimensional image as follows:

For each pixel, a line strength measure (S) is calculated by applying

$$S = (L - N) \tag{4.1}$$

in 12 equally–spaced orientations, where L is the mean grey–level value of a line of pixels of length 5 centred on the target pixel, and N is the mean grey–level value of a similarly–orientated square of pixels. For each pixel, the maximum S is selected and its corresponding orientation taken to be the pixel's line orientation.

This method was adapted for use on the three-dimensional tomosynthesis images by calculating the mean grey–level of a line of 5 voxels at a range of 12 orientations for each of 12 planes orientated around the x–axis, and subtracting the mean grey–level of a similarly–orientated cube of voxels. Voxel values were obtained using trilinear interpolation where positions did not correspond with individual voxel positions. Prior to processing, each slice was scaled to produce a pixel resolution approximately equal to the between–slice resolution.

As with the 2–dimensional method, a multi-scale approach was used in order

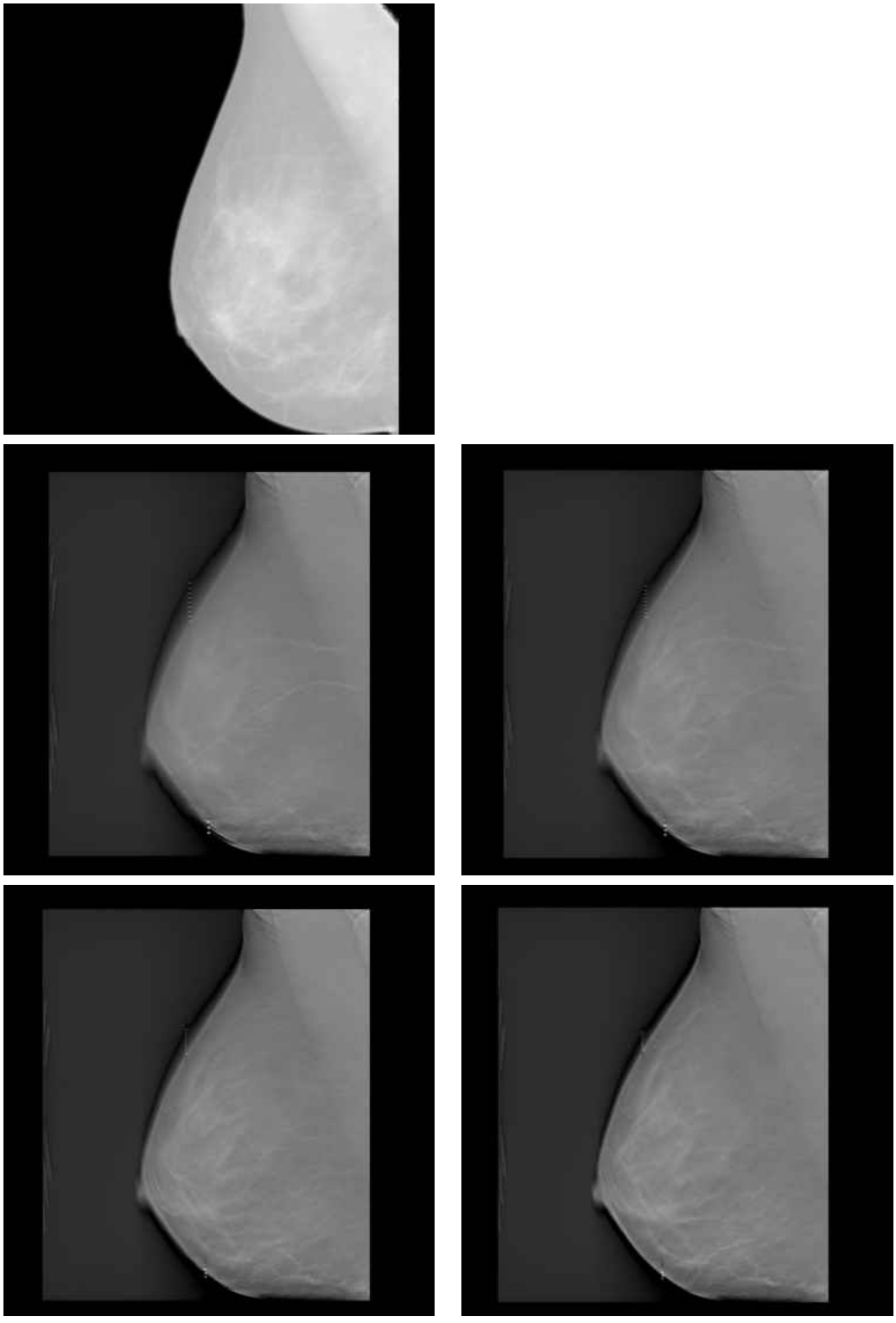


Figure 4.7: A typical mammogram (top left) and four slices from the corresponding tomosynthesis volume. The first slice (middle left) is from close to the edge of the volume and shows the skin layer, whereas the fourth slice (bottom right) is from close to the centre of the volume and shows the deeper fibro-glandular tissue.

to detect lines of a range of thicknesses whereby the images were blurred using a 3x3x3 Gaussian filter then subsampled to produce an image at half the width and height of the original. A Gaussian of 3x3x3 ($\sigma = 0.85$) was selected because it was felt that a larger Gaussian may blur the image too much and result in a loss of clarity which may lead to poorer results from the subsequent linear structure detection. The line detector was applied independently on the images at each scale and then combined by taking the maximum line strength value from the corresponding voxel(s) in each scaled image. Three scales were used with the 2D mammograms and two scales with the tomosynthesis images, since the original tomosynthesis images were smaller than the original mammograms.

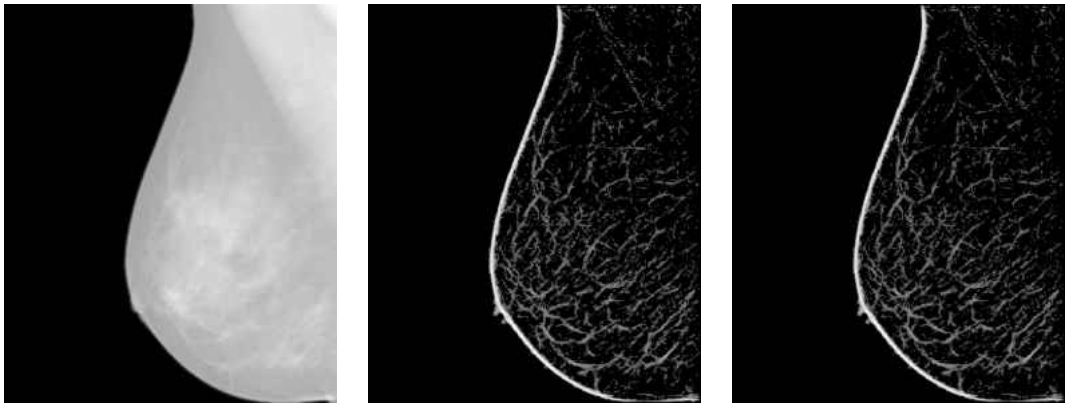
Finally, the voxel line strengths were thresholded to remove background texture and artefacts. A measure of above–threshold linearity was calculated for each image as an indicator of linear density. Areas outside of the breast area were masked and discarded.

Linearity measures for corresponding 2D and 3D images for each patient were compared.

4.3.4 Results

Figure 4.8 shows line detection results for a typical two–dimensional mammogram. The figure shows the original digital projection mammogram alongside the unthresholded and thresholded line detection result images.

Typical line detection results for the three–dimensional tomosynthesis images are shown in Fig. 4.9. This figure shows sample results for a series of slices in 5 slice increments. For each slice, the original slice image is displayed alongside both the unthresholded and thresholded line detection result images. The initial output images are thresholded at a level most suitable for removing background and



(a) (b) (c)

Figure 4.8: Mammogram line detection results. Original mammogram (a) alongside the unthresholded line detection results (b) and thresholded line detection results (c).

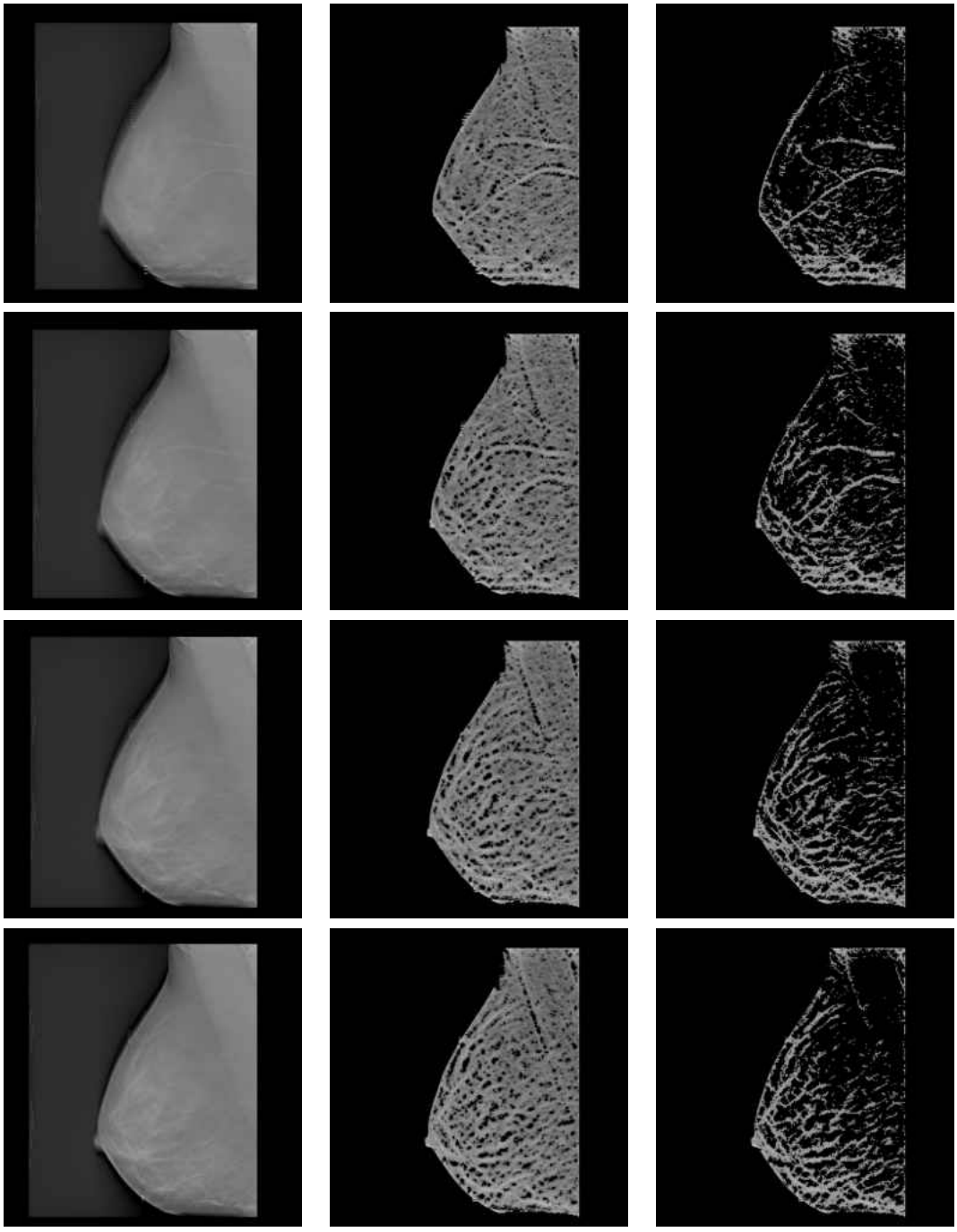
reconstruction artefact noise whilst maintaining the linear structure information.

Figure 4.10 shows a graph of the detected linearity in two–dimensional mammograms against the detected linearity each corresponding three–dimensional tomosynthesis image. The results demonstrate a good degree of correlation (Pearson's coefficient 0.73).

4.3.5 Discussion

The comparison between the linearity detected in two–dimensional mammograms and corresponding three–dimensional tomosynthesis images shows a good degree of correlation, achieving a correlation coefficient of 0.73. Since the linearity in two–dimensional mammograms has been well–correlated to mammographic risk [106, 107], it can be expected that the linearity in tomosynthesis images may also be correlated to risk.

This study is limited in several aspects by the lack of availability of a wide range of image data. At the time of these experiments, breast tomosynthesis is experimental and only a handful of tomosynthesis machines currently exist. As such most of the images taken are of patients with known cancers during clinical



(a) (b) (c)

Figure 4.9: Tomosynthesis line detection results. Each row shows an image slice with an interval of 5 slices between rows. The top row shows a slice close to the edge of the volume, whilst the bottom row shows a slice close to the centre of the volume. In each row, column (a) shows the original tomosynthesis slice, column (b) shows the unthresholded line detection results and column (c) shows the thresholded line detection results.

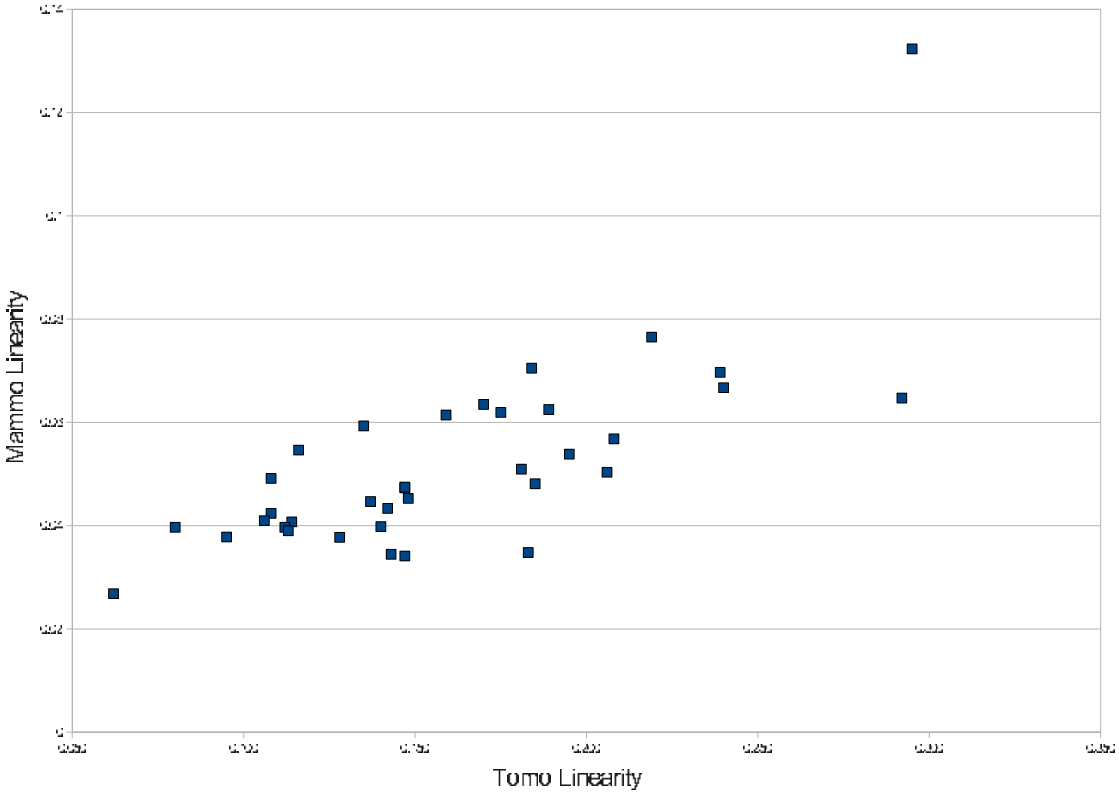


Figure 4.10: A comparison of the linearity of two–dimensional mammograms and corresponding three–dimensional tomosynthesis images.

trials [57]. The small dataset used (39 images) is clearly a limitation of this study.

A direct comparison between three–dimensional linearity and mammographic risk would be an ideal continuation, however this would be difficult in practice due to the limited data available. Such a comparison would likely require a large dataset of tomosynthesis images that are spread across all risk categories, whereas currently–available datasets are generally small and images are concentrated around the high risk categories, making any valid statistical comparison unlikely.

Other recent studies have suggested a link between the parenchymal pattern visible in tomosynthesis images and mammographic risk [108], and have shown that the parenchymal patterns are more clearly defined in tomosynthesis volumes [60]. It was also suggested that, due to the advantages of three–dimensional images in mammography, the parenchymal patterns visible in three–dimensional tomosyn-

thesis images might eventually be expected to provide a better indicator of risk than those currently available from two–dimensional mammograms [108].

Therefore, a perfect correlation result in this experiment would not be ideal, as it would indicate that there is no additional information available in the three–dimensional tomosynthesis images.

In summary, this experiment has demonstrated that linear structures can be detected in 3–dimensional tomosynthesis volumes. This can be used as a basis for further work to produce a risk classifier.

4.4 Network Representation

4.4.1 Introduction

It has been demonstrated that mammographic risk is correlated to the density of linear structures (see Chapter 3), however it has been suggested that risk may also be related to the distribution of those structures [21, 22]. In order to further study the distribution of linear structures in mammographic images a 3–dimensional graph–like network representation of the linear structure distribution would be beneficial. A network representation would allow for factors related to the distribution of linear structures to be measured, such as the length and number of edges and nodes. The development of this method would be beneficial over a more statistical approach because with further study it has the potential for greater use beyond the scope of this project, including areas such as abnormality detection.

4.4.2 Method

The complete process of the detection of linear structures and the production of the 3D network representation can be summarised as follows:

1. A background mask is produced to identify the non–breast area
2. The image volume is scaled by blurring and subsampling the image
3. The Line Operator method is performed, providing a resulting line strength image
4. The non–breast area is removed by applying the mask produced in stage 1
5. The line strength image is thresholded using the double–threshold approach
6. The directional Z–profile non–maximum suppression method is applied to reduce the effects of distortions and artefacts
7. Nodes are identified by locating line strength maxima
8. The line strength image is merged with the unenhanced, unthresholded image produced in step 3
9. The A* search algorithm is used to connect neighbouring nodes, producing a 3D network of nodes and edges

The development of this method is discussed in this section.

4.4.2.1 Network Extraction

Left and right 3–dimensional tomosynthesis images were captured and reconstructed for 39 women (see Sec. 4.3.2). The line operator method described in Sec. 4.3.3 was performed on each image resulting in a measure of line strength for

each voxel in the 3D volume. A mask was applied to remove parts of the image outside of the breast area, and the resulting line strength values were thresholded to remove low–level responses from background tissue.

The network was produced in several stages. Firstly, nodes were identified in the 3D line strength volume. It was discovered that the linear structures in the line strength volume were not sufficiently well–defined for junction detection to be effective. Instead, line strength maxima (voxels where the line strength values were greater than all surrounding voxels) were identified and used as initial nodes. Since the network representation was intended to provide an insight into the relative distributions of linear structures it was felt that the line strength maxima would make suitable nodes since the resulting effects at junctions would be similar in all images.

For each node, the 8 nearest nodes by direct Euclidean distance were identified, and the edges connecting each of these nearest nodes with the target node were added to a map of potential edges. An A* search was used to attempt to join each of the pairs of nodes along each potential edge. The search path was constrained such that the path could not pass through any voxel with a line strength below the threshold, or outside of the masked breast area. In addition, any path that exceeded $1.5 \times d$, where d is the direct Euclidean distance between the start and end points, was discarded in order to improve performance by limiting the search space for each potential edge.

The line strength was given a penalty in order to discourage the search path from straying away from high–line strength areas. The purpose of this was to encourage the search path to follow to the linear structures as closely as possible. The path distance PD between starting voxel V_0 and a given voxel V_p along the

path is given by

$$PD\left(V_0, V_p\right) = PD\left(V_0, V_{p-1}\right) + d\left(V_{p-1}, V_p\right) + \left(\frac{L\left(V_0\right) + L\left(V_n\right)}{2} - L\left(V_p\right)\right) \quad (4.2)$$

where d is the direct Euclidean distance between two voxels, L is the line strength value for a given voxel and V_n is the ending voxel.

This line strength penalty permitted the path to enter low–line strength areas (subject to the threshold) where necessary, but encouraged this to remain limited and to seek a high–response path if possible. This is important because the level of the line strength response fluctuated along the linear structures due to many factors including the type of surrounding tissue (and hence the contrast between the linear structure and the background), capture and reconstruction and artefacts and junctions and other variations of the linear structures themselves. Figure 4.11 shows an example of a linear structure containing an area of low line strength response between two areas of high response.

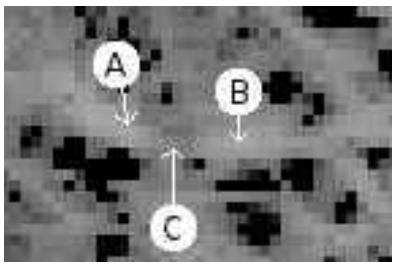


Figure 4.11: An example showing a linear structure with two nodes (A & B) and an area of low line strength response (C) between them.

A 3–dimensional network was constructed using the nodes identified from the line strength volume, joined by edges where connections could be made using the A* approach. All nodes that had no edges connecting to them were assumed to be background noise and discarded. Due to the approach of using line strength maxima as nodes and fluctuations in the line strength responses due to background tissue, other linear structures, reconstruction artefacts and other factors, a typical linear structure could feature many nodes along its length. In order to minimise

this effect, any node joined by exactly two edges was removed and the two edges combined to form a single edge (see Fig. 4.12).

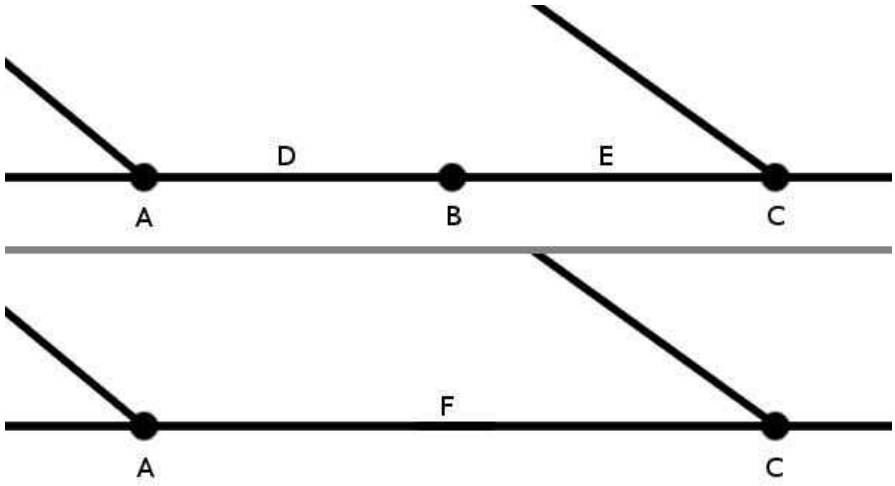


Figure 4.12: An example of a node (B) connected by exactly two edges (D & E). Node B is removed and edges D & E are replaced by a single edge (F) connecting nodes A & C.

4.4.2.2 Initial Network Results

Figures 4.13–4.14 show the results of the network extraction algorithm described in Sec. 4.4.2.1 applied to the image volume shown in Fig. 4.9.

These initial results were quite poor because the purpose of this experiment was to create a realistic network representation of the network of linear structures, however the network results show little relation to the linear structures visible in the line detection results (see Fig. 4.9). Typically, linear structures were not represented in the network, whilst the network contained many erroneous edges that could not be related to visible linear structures in the line detection results.

It is believed that two principal factors have contributed to the poor results demonstrated. The first of these is that breaks often occur in linear structures where the line strength response has dipped below the threshold used to remove background tissue, and as a result the A* search algorithm cannot complete a

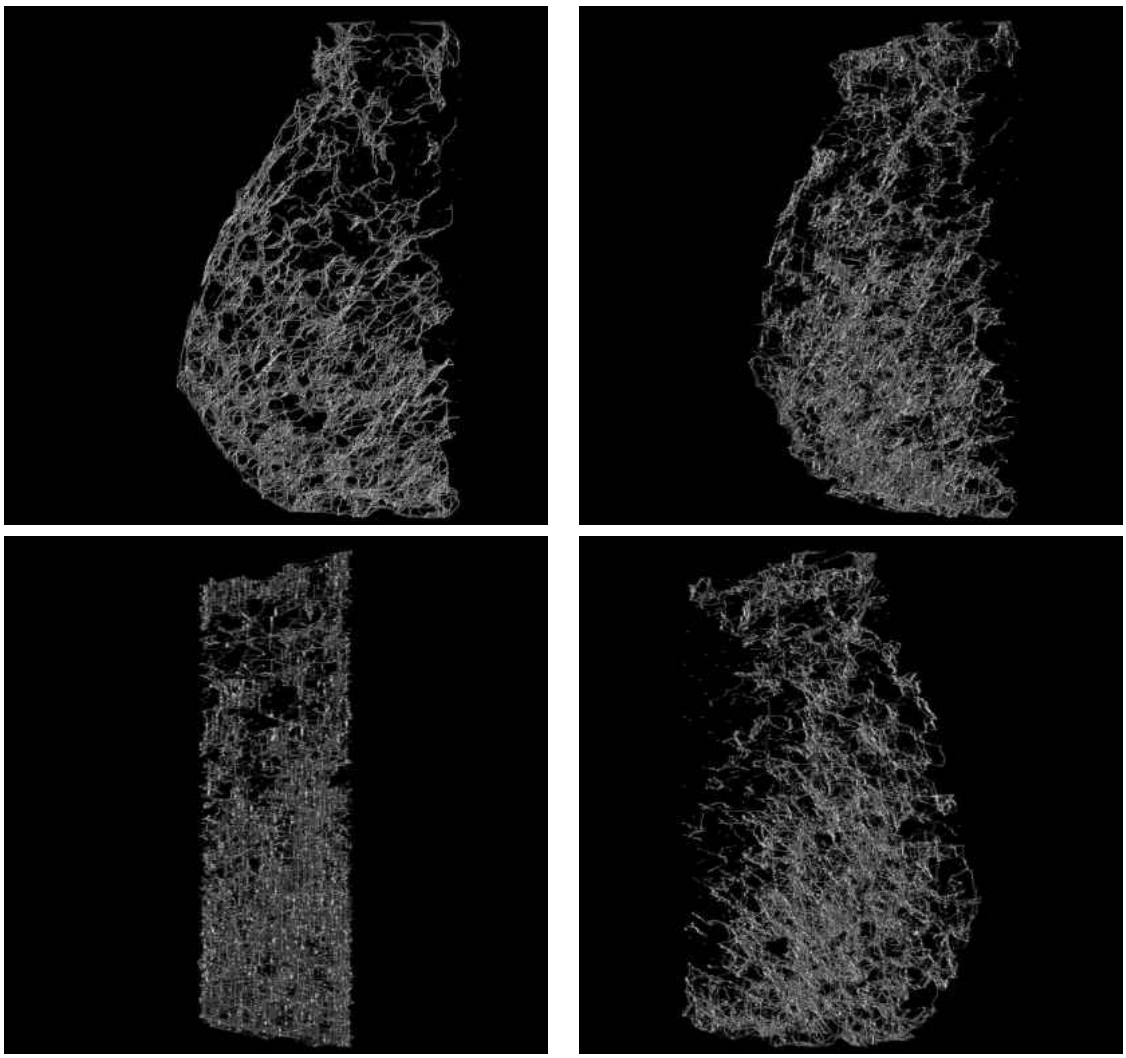


Figure 4.13: Rendering of a typical 3–dimensional network shown at various orientations (continued in Fig. 4.14).

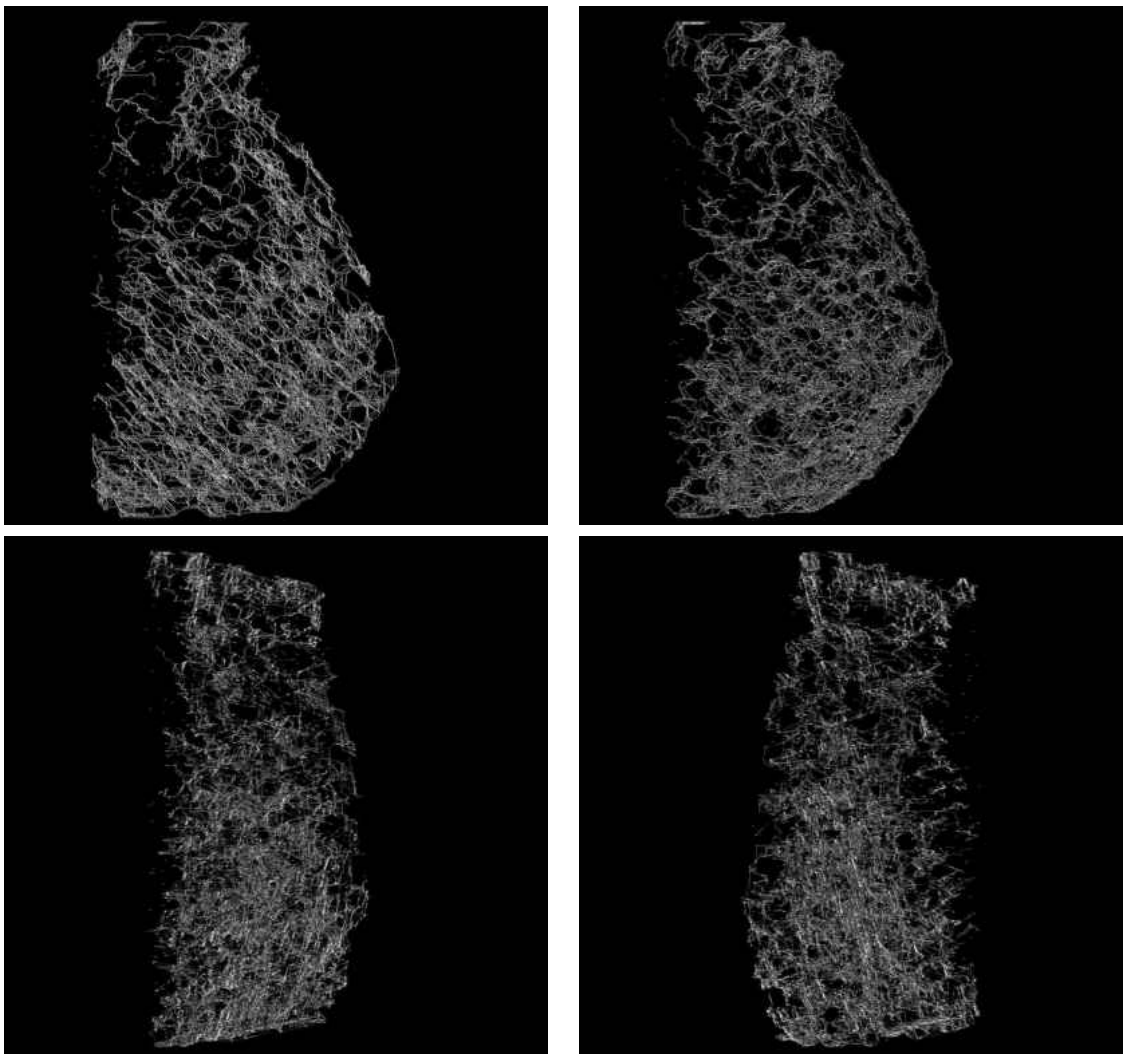


Figure 4.14: Rendering of a typical 3–dimensional network shown at various orientations (continued from Fig. 4.13).

link between two nodes where a linear structure exists. The second factor is the considerable distortions and artefacts present in the image as a result of the reconstruction and image capture, which may be a result of perorming 3D resonstruction from a small number of images across a relatively small arc.

4.4.2.3 Artefacts and Distortions

A significant problem encountered during the network extraction process was the widespread presence of distortions and artefacts in the reconstructed tomosynthesis volumes. This was probably due to the difficulty of reconstructing a complete 3D volume from only a small number of raw images across a narrow arc. Typically, linear and other structures within the breast appear 'stretched' in the Z–axis such that they are visible across many more image slices than would be correct for their actual dimensions. An example of this effect is show in Fig. 4.15, where the highlighted linear structure is becoming visible in the first slice (9mm), is strongly visible in the slices at 15mm and 20mm, and is finally fading in the slice at 24mm. This would indicate that the structure is approximately 15mm wide, which is unlikely to be true.

Figure 4.16 shows a graph of the Z–axis grey–scale profile of a similar distorted linear structure, showing its visibility through the image volume. It may be hypothesised that this effect could be beneficial for some applications (such as abnormality detection), however it makes the process of extracting a network that is representative of the linear structure distribution very difficult.

Two principal factors are known to affect the magnitude of the distortions in the reconstructed tomosynthesis volumes: the overall acquisition angle of the raw image sequence and the reconstruction algorithm used [109, 110]. Whilst the simultaneous algebraic reconstruction technique (SART) has been shown to perform better than the filtered–backprojection approach [109], the overall acquisition

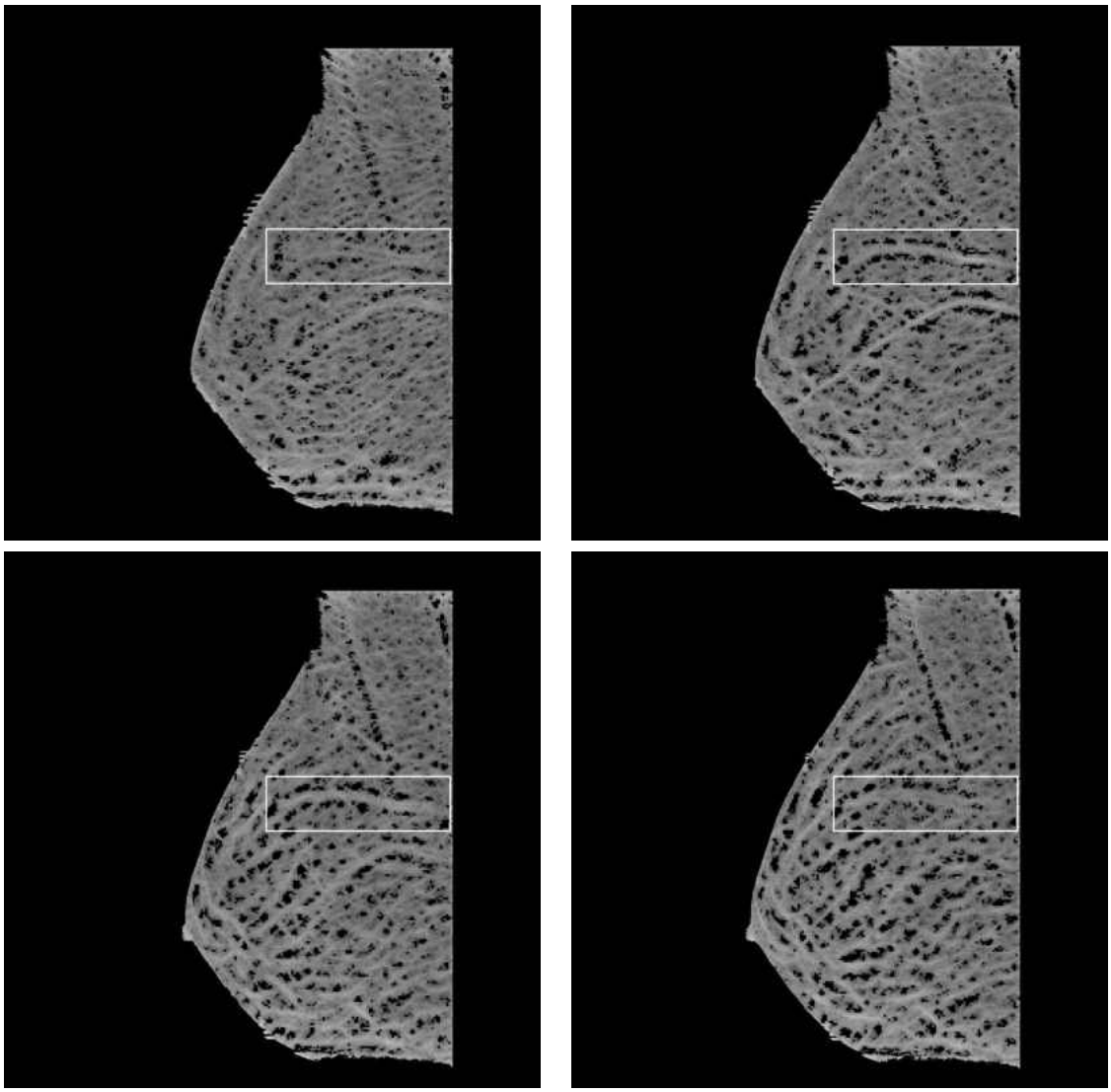


Figure 4.15: Slices from the line operator results for typical mammogram highlighting a distorted linear structure (indicated by white rectangle). Slices from top left to bottom right are at 9mm, 15mm, 20mm and 24mm.

angle is the predominant effect on the level of interslice blurring [109, 110].

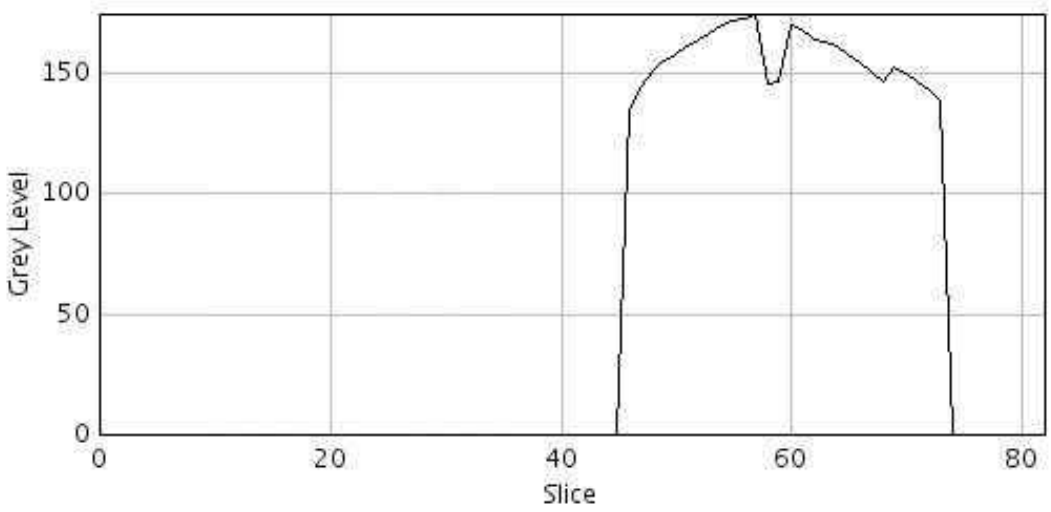


Figure 4.16: A graph showing the Z–axis grey–scale profile of a distorted linear structure.

Initial attempts to reduce the effects of distortions revolved around the application of a line thinning algorithm. An algorithm based on one described by Pálágyi and Kuba [111] was implemented and applied to the line operator result volumes. The selected algorithm was chosen because it reduces 3–dimensional shapes to a medial skeleton line structure without creating a medial surface. It is also straightforward to implement, efficient for use with datasets of large images and performs well in noisy cases [111].

The thinning algorithm defines a set of twelve $3 \times 3 \times 3$ deletion masks. Voxels within each mask may be defined as either linear structure or background, such that each of the masks will match border voxels in a particular direction. Border voxels are voxels which are classified as belonging to a linear structure but where one or more neighbouring voxel is classified as background.

During each iteration of the algorithm, the twelve deletion masks are applied to each linear structure voxel in the image and the voxel removed (changed to background) if at least one mask matches its $3 \times 3 \times 3$ neighbourhood. The algorithm continues until no further border voxels are removed.

The algorithm was applied to the thresholded line detection results (see Fig. 4.17).

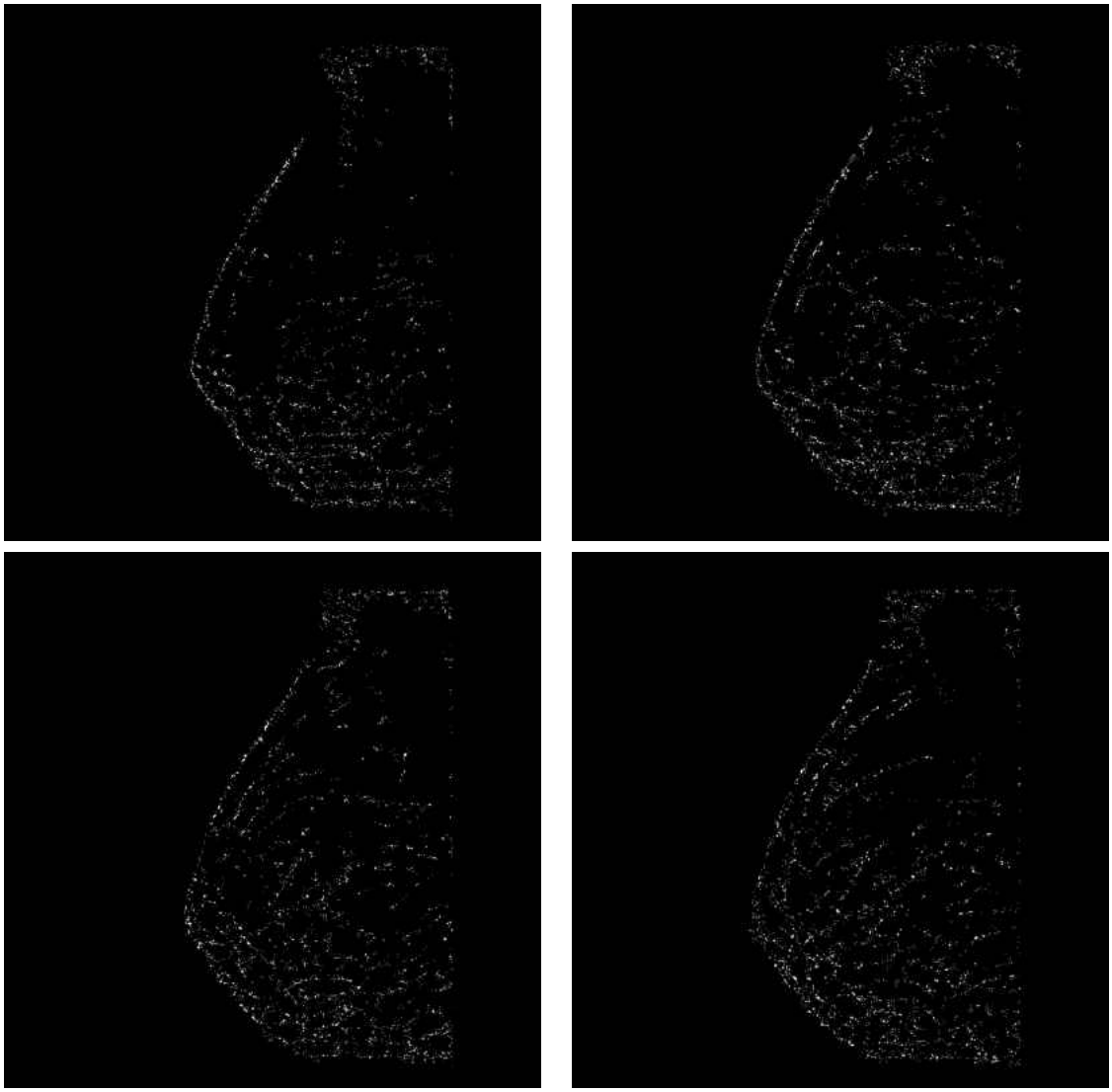


Figure 4.17: The slices previously shown in Fig. 4.15 after processing the line operator results using the line thinning algorithm. Slices from top left to bottom right are at 9mm, 15mm, 20mm and 24mm. Images have been enhanced to improve visualisation.

The results in Fig. 4.17 show that the proportion of voxels classified as linear structures has been reduced considerably, with the skeleton lines difficult to trace when viewing individual slices of the image volume. However, two principal problems were identified which ultimately led to the approach being deemed unsuitable for this application.

The first problem discovered when using this approach was that the previously–mentioned problem of breaks occurring in linear structures where the line strength

dipped below the background threshold (see Sec. 4.4.2.2) was exacerbated by the thinning algorithm. Where a small break occurred within a broad linear structure, the border voxels were removed not just along the side edges of the linear structure but also at the 'broken' ends, thus increasing the size of the break. This effect can be seen in Fig. 4.18. Although at this stage the principal factor affecting the ability of the A* search in the network extraction method to trace along a linear structure is simply the presence of breaks rather than their size, the large breaks made efforts to overcome these breaks considerably more difficult.

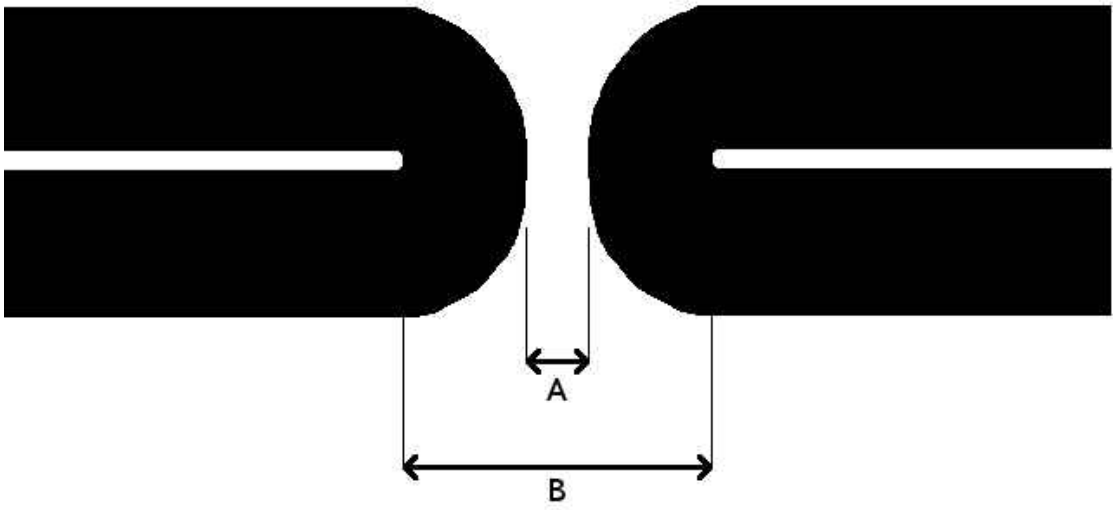


Figure 4.18: A diagram showing a broad linear structure (black) with a break (A). Following application of the thinning algorithm the broad linear structure is reduced to the skeletal lines (white) separated by a larger break (B).

Problems with the thinning algorithm were also present where either the distortions in the Z–axis were very large relative to the overall dimensions of the linear structure or where (often due to distortions in the Z–axis) different linear structures crossed. It was discovered that a short section of linear structure may be 'stretched' in the Z–axis such that it becomes longer in the Z–axis than in its true direction. As a result, the medial line produced by the thinning algorithm will follow the direction of the 'stretching' in the Z–axis rather than the true direction of the linear structure. This effect is illustrated in Fig. 4.19.

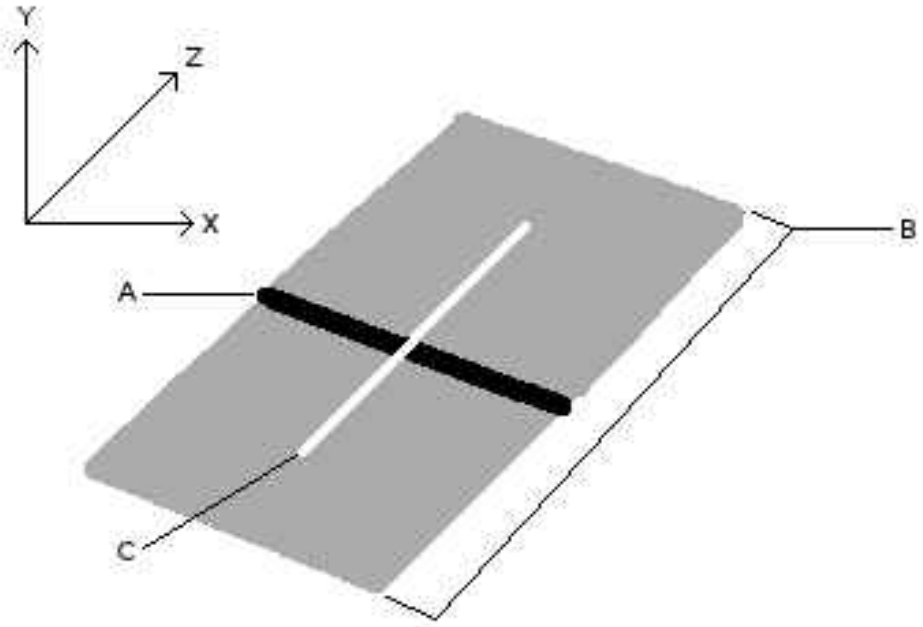


Figure 4.19: A diagram showing a short linear structure (A). The linear structure visible in the reconstructed tomosynthesis volume has been stretched in the Z–axis by more than the length of the linear structure (B). This results in the thinning algorithm producing a medial line (C) approximately perpendicular to the original linear structure.

The thinning algorithm also produced anomalous results in cases where two linear structures that do not intersect, when stretched in the Z–axis do intersect. This led the thinning algorithm to view the two structures as a single 3D structure and extract a medial line for the entire structure. This medial line would inevitably bear little resemblance to the two original linear structures.

These limitations and anomalous results led to the abandonment of the thinning algorithm and the development of an alternative approach. It was realised that a thinning algorithm that attempted to reduce the 3–dimensional structure in all three dimensions was not desirable since the distortions only occurred in the Z–axis, and the X and Y–axis information was correct. The devised approach was to use a form of non–maximal suppression (NMS) in the Z–axis.

For each voxel in the $X-Y$ plane a grey–scale profile of voxels in the Z axis was

examined. A typical Z–profile is shown in Fig. 4.16, showing a 'stretched' linear structure. The NMS approach was performed on the thresholded line strength volumes, as such there are no voxels with line strength values below a threshold value.

Each Z-profile was smoothed using a 7–voxel Gaussian mask to reduce the effects of local maxima and minima. Individual structures were identified by locating continuous sections of the Z–profile that were above zero (i.e. above the line strength threshold, and so had not been removed during thresholding). For each structure the median voxel was located and the structure was reduced to a maximum of 5 voxels centred on the median voxel. All voxels that were part of the structure but not within this central section were given a grey–level value of zero (background tissue).

The result of performing this algorithm on the Z–profile shown in Fig. 4.16 is shown in Fig. 4.20. The 'stretched' structure has been reduced to a narrow core around its centre.

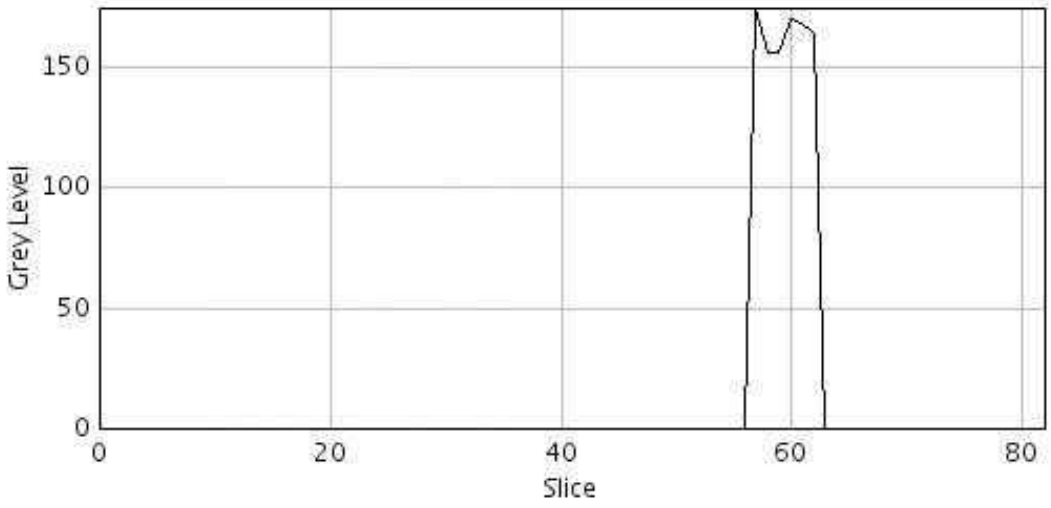


Figure 4.20: A graph showing the Z–axis grey–scale profile of a distorted linear structure as shown in Fig. 4.16 after processing the line operator results using the Z–profile NMS approach.

Further investigation revealed that although the Z-profile had been success-

fully reduced to a narrow core, the images still displayed signs of blurring and 'stretching' around the visible structures in the images. It was discovered that this was due to the shape of the reconstructed data within the image volumes. It was assumed that the breast data was in a cubic shape within the image volume, i.e. the reconstructed breast data would be the same size across all slices, however this was not correct. Due to the image capture geometry the breast data was trapezoid in shape when viewed across the slices, with the area of the reconstructed breast data in each slice becoming progressively larger through the image volume as if projected from a single point. This is illustrated in Fig. 4.21.

It was suggested that since the reconstruction was not cubic, the 'stretching' of structures between slices was not occurring in a line parallel to the Z axis but was occurring in a line directed at the 'projection origin'. It was therefore proposed that the Z-profile non-maximum suppression be modified such that instead of using profiles parallel with the Z axis, the profiles should be in the direction of the 'projection origin'.

For each image volume, the 'projection origin' point was calculated by measuring the maximum extent of the reconstructed data area of each slice and projecting a line along the corners of this area until they intersected. For each pixel in the outermost slice, a line of pixels was extracted along a path in the direction of the origin point and the same Z-profile NMS algorithm described above was performed. Where a pixel was extracted in more than one line, the maximum value was used as its final value.

The resulting images (examples shown in Fig. 4.22) demonstrated a significantly reduced level of blurring and stretching artefacts.

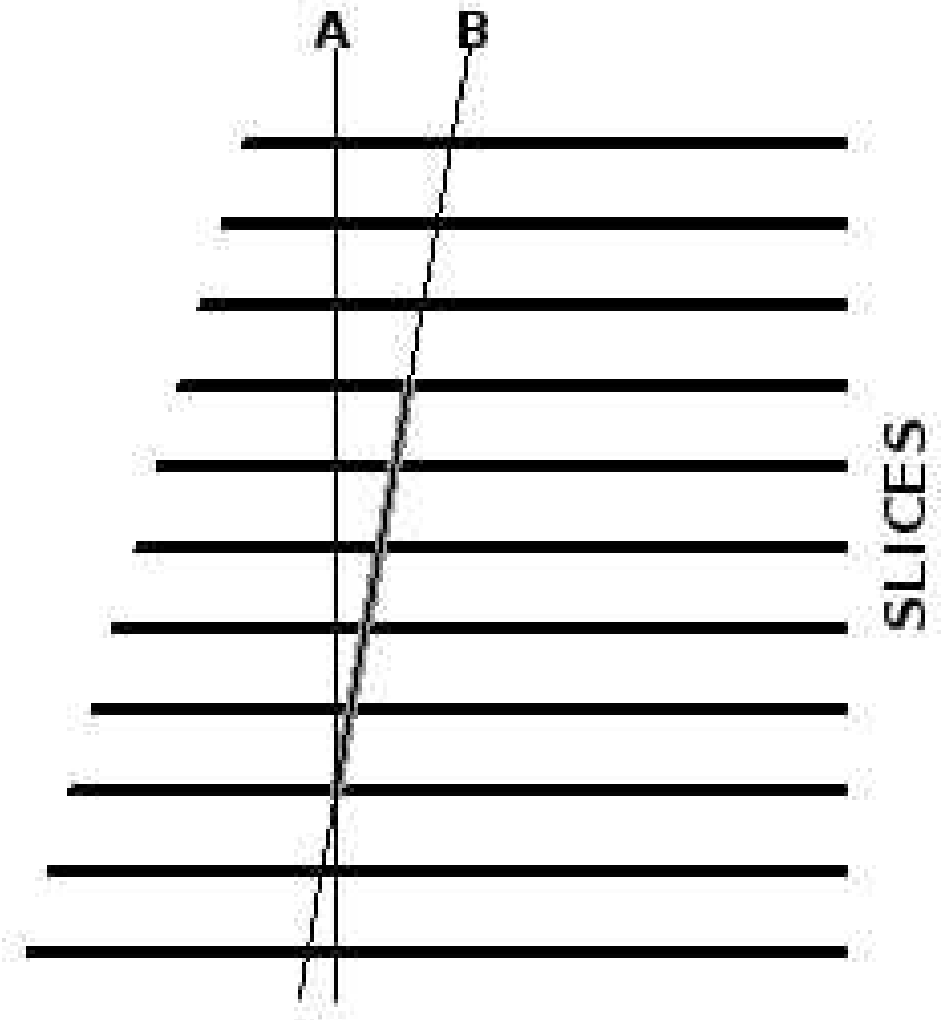


Figure 4.21: An illustration of the reconstructed data within a set of image slices. A shows a Z-profile parallel to the Z axis, whereas B shows a Z profile directed at the projection origin.

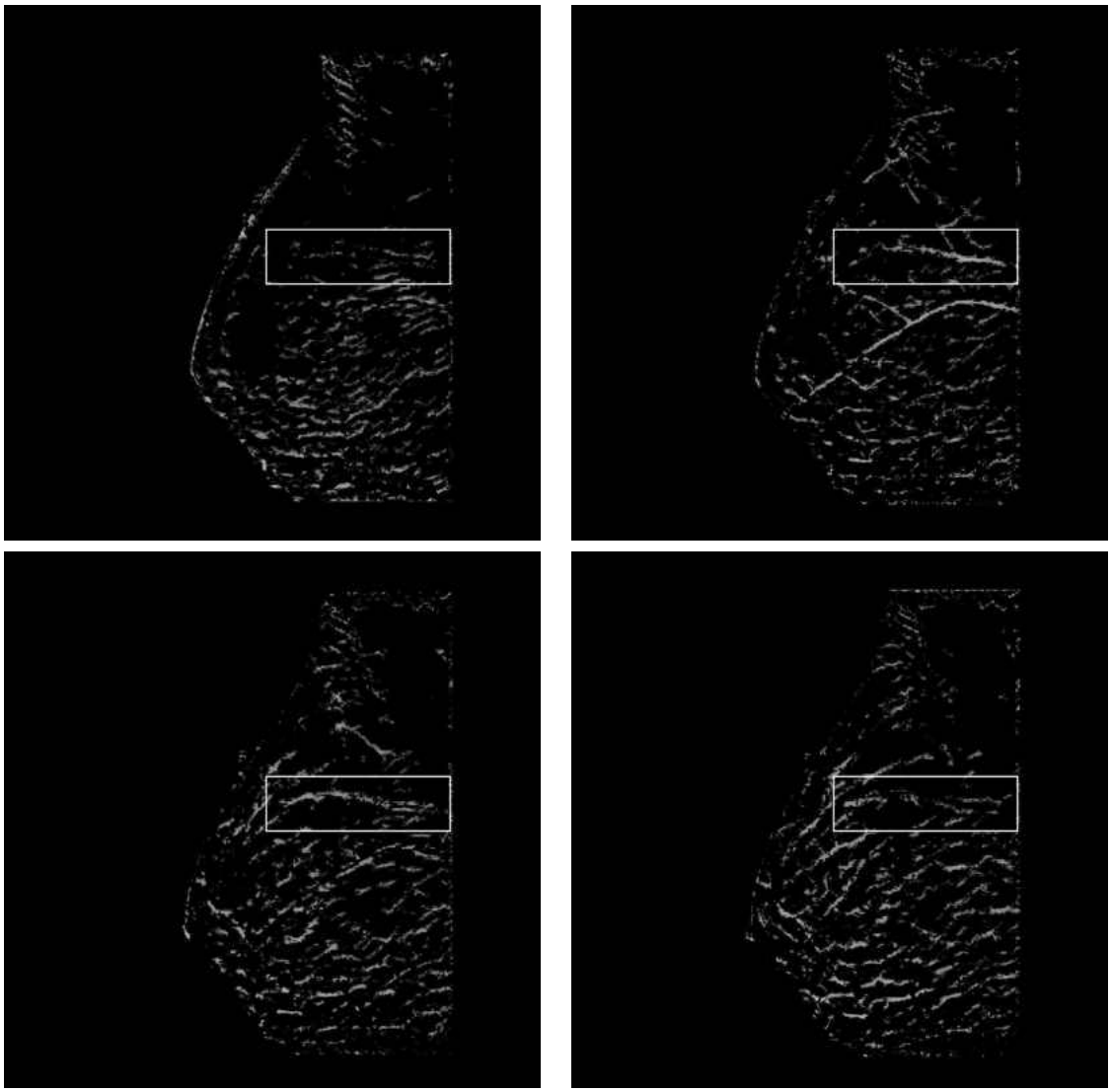


Figure 4.22: The slices previously shown in Fig. 4.15 after processing the line operator results using the Z–profile NMS approach. Slices from top left to bottom right are at 9mm, 15mm, 20mm and 24mm.

4.4.2.4 Areas of Low Line Strength in Linear Structures

Once the problem of artefacts and reconstruction distortions had been addressed, the remaining problem identified (see Sec. 4.4.2.2) was the issue of 'gaps' in linear structures. Various factors could lead to the line strength values in localised areas reducing significantly. Such factors include dense tissue (and hence low contrast) in the surrounding area and intersecting or nearby lines. These caused gaps to appear in the thresholded line strength images where the line strength periodically dropped below the threshold, leading to the failure of the A* search algorithm to follow the linear structure correctly.

A simple reduction in the threshold was deemed to be unsuitable since the background noise would no longer be suitably removed, so a more complex thresholding approach was used. The technique developed was modelled on that used in the Canny Edge Detector [112], whereby two thresholds are used.

Previously a single threshold T was be used, where all voxels of line strength greater than or equal to T would be included whilst all voxels of line strength less than T would be excluded. This was replaced by both an upper threshold T_U and a lower threshold T_L. All voxels of line strength greater than or equal to T_U would be included, whilst all voxels of line strength less than T_L would be excluded. Voxels of line strength greater than or equal to T_L but less than T_U were included only if they were connected, directly or indirectly, to a voxel of of line strength equal or greater than T_U (see Fig. 4.23).

The described approach has the advantage of permitting a low threshold within linear structures, where the line strength will invariably rise above T_U, whilst continuing to exclude background noise.

The principal disadvantage of this approach is its computational complexity. Whilst it is easy to identify all voxels with line strength above a particular thresh-

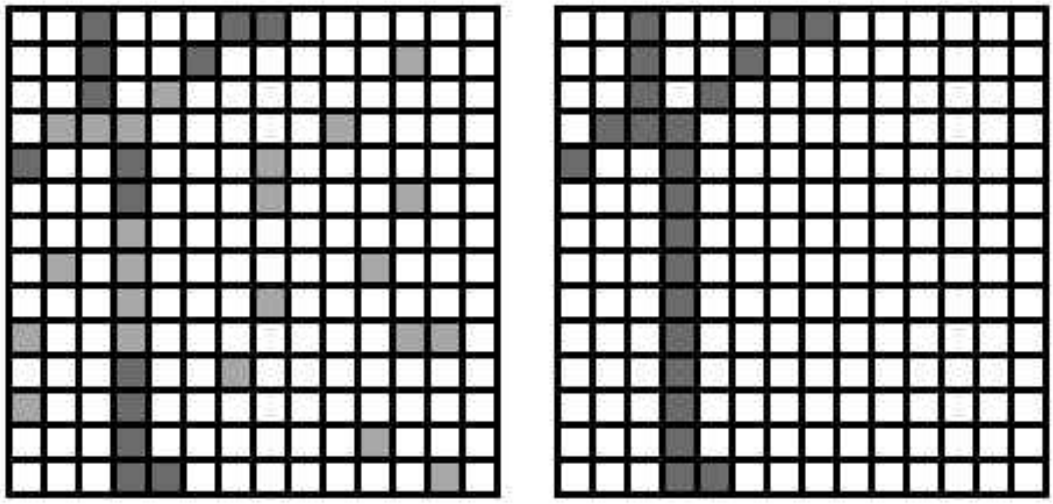


Figure 4.23: An illustration of the effects of the bi–threshold approach. In the left image, dark grey squares represent pixels of line strength greater than or equal to T_U and light grey squares represent pixels of line strength greater than or equal to T_L but less than T_U. Light grey squares include several pixels with linear structures and several background pixels. The right image shows the result after processing, showing that the light grey squares that were connected to dark grey squares have been included whereas isolated light grey squares have been excluded.

old, or even between two thresholds, it is more time consuming to identify only those connected to voxels above the higher threshold. An iterative approach was used whereby all voxels of line strength greater than or equal to T_U were identified and included in the resulting image, we shall call these included voxels V_I, and all voxels of line strength greater than or equal to T_L but less than T_U were identified, we shall call these candidate voxels V_C.

During the course of a single iteration, each voxel in V_I was visited its 26 neighbouring voxels were checked. Any of the neighbouring voxels that had been marked as candidate voxels V_C were added to V_I. During the next iteration, the voxels that were added to V_I in the previous iteration were visited and their neighbouring voxels were checked. This continued until no voxels were added to V_I.

Example results are shown in Figs. 4.24–4.25. The images clearly show the linear structures in a more pronounced and more complete state when using the

double thresholding approach.

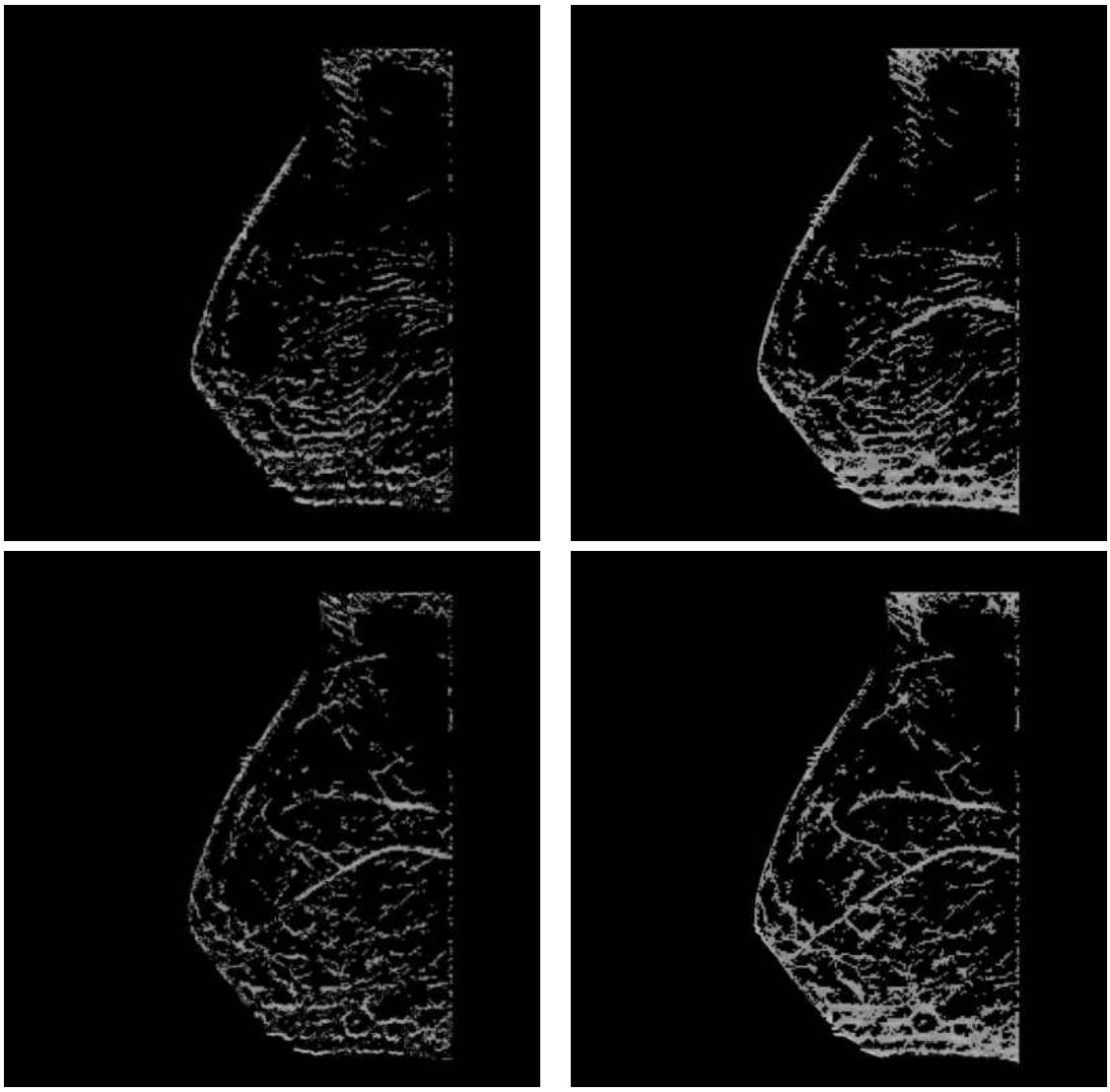


Figure 4.24: Images on the left are those previously seen in Fig. 4.22 and show a selection of slices of Line Operator results that have been subjected to standard thresholding and processing using the Z–Profile NMS approach (see Sec. 4.4.2.3). Images on the right show the same slices where the double thresholding approach has been used, followed by processing with the Z–Profile NMS technique. Slices at top are at 9mm and at bottom are at 15mm. Slices at 20mm and 24mm are shown in Fig. 4.25.

Although the new approach resulted in fewer breaks in the linear structures than the previous approach, breaks remained in the thresholded image that caused the A* search to fail to connect nodes that ought to have been connected. Reducing the minimum threshold further was not deemed to be a viable option due to the inclusion of background noise resulting in spurious 'structures' in the network.

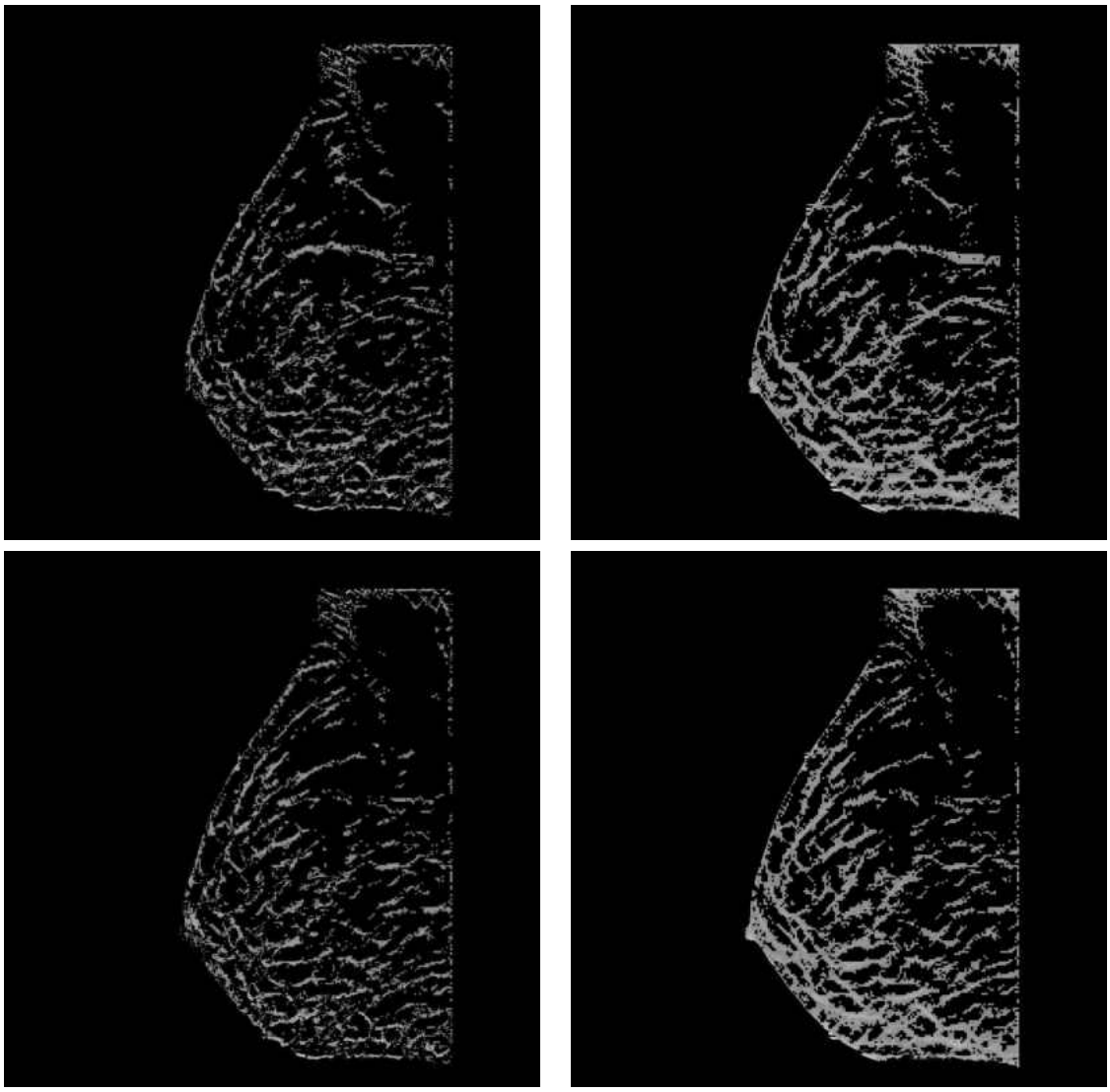


Figure 4.25: Continued from Fig. 4.24. Images on the left are those previously seen in Fig. 4.22 and show a selection of slices of Line Operator results that have been subjected to standard thresholding and processing using the Z–Profile NMS approach (see Sec. 4.4.2.3). Images on the right show the same slices where the double thresholding approach has been used, followed by processing with the Z–Profile NMS technique. Slices at top are at 20mm and at bottom are at 24mm. Slices at 9mm and 15mm were shown in Fig. 4.24.

A solution was devised that involved merging the thresholded image with the unenhanced, unthresholded Line Operator results image at a stage after network nodes have been identified but before the A* search is performed. The merge was performed on a voxel–by–voxel basis subject to the following criteria:

- Voxels outside of the masked breast area are discarded,
- Voxels classified as part of a linear structure in the thresholded image use their grey–scale value from the thresholded image,
- Voxels not classified as part of a linear structure in the thresholded image use their grey–scale value from the unenhanced, unthresholded image subject to a multiplier (0.5).

The effect of performing the A* search on the merged image was that small gaps in a linear structure could be overcome, although this would be at a penalty to the search algorithm.

Figures 4.26–4.27 show a 3D rendering of a typical network structure produced using the full enhanced process described.

4.5 Classification

4.5.1 Introduction

The purpose of this experiment is to investigate how the distribution of linear structures might be related to mammographic risk. The method described in Sec. 4.4 is applied to extract a network representation for each of a set of tomosynthesis images, and data from these networks relating to the distribution of linear structures is used to classify the images according to mammographic risk.

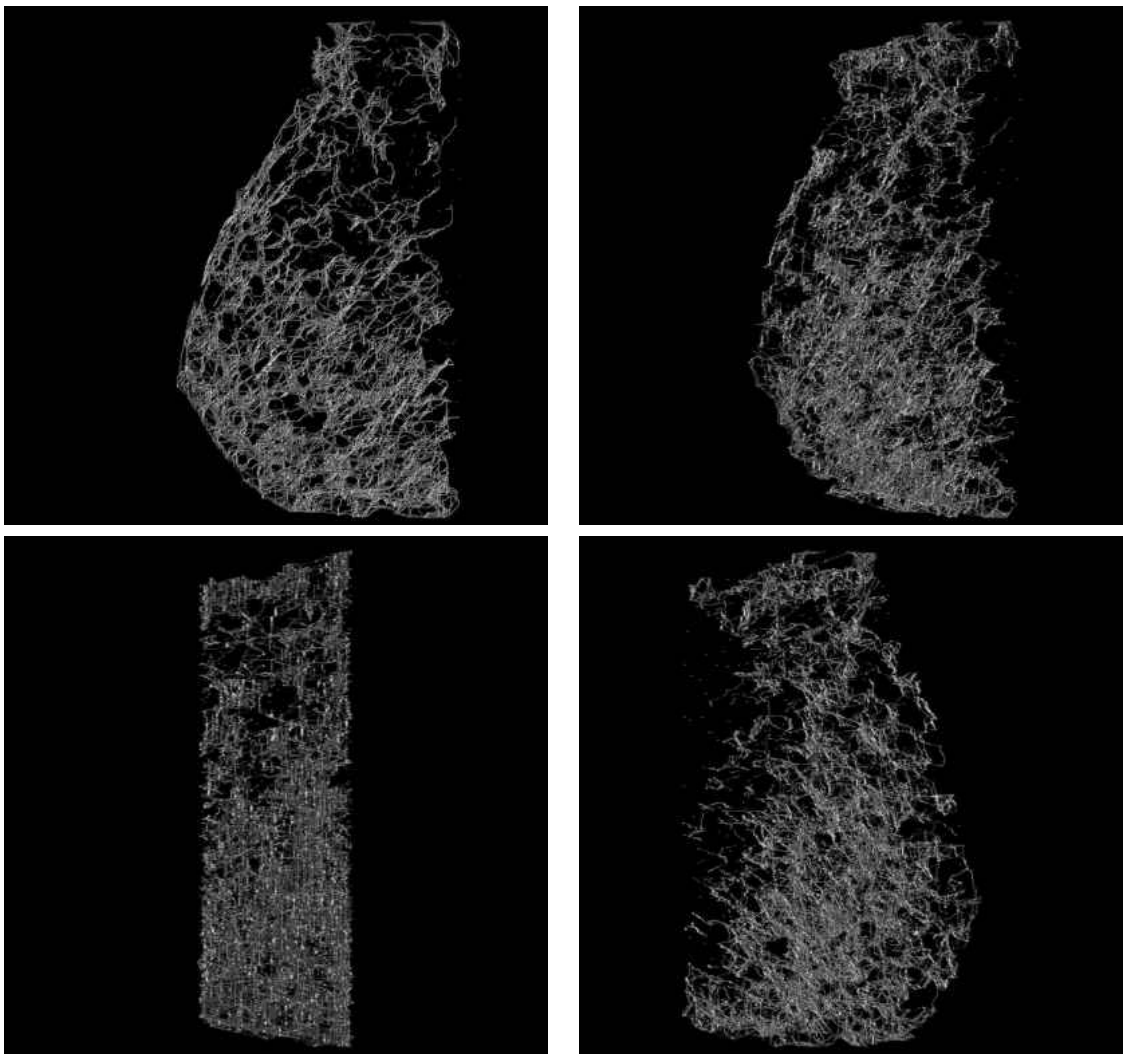


Figure 4.26: Rendering of a typical 3–dimensional network shown at various orientations (continued in Fig. 4.27).

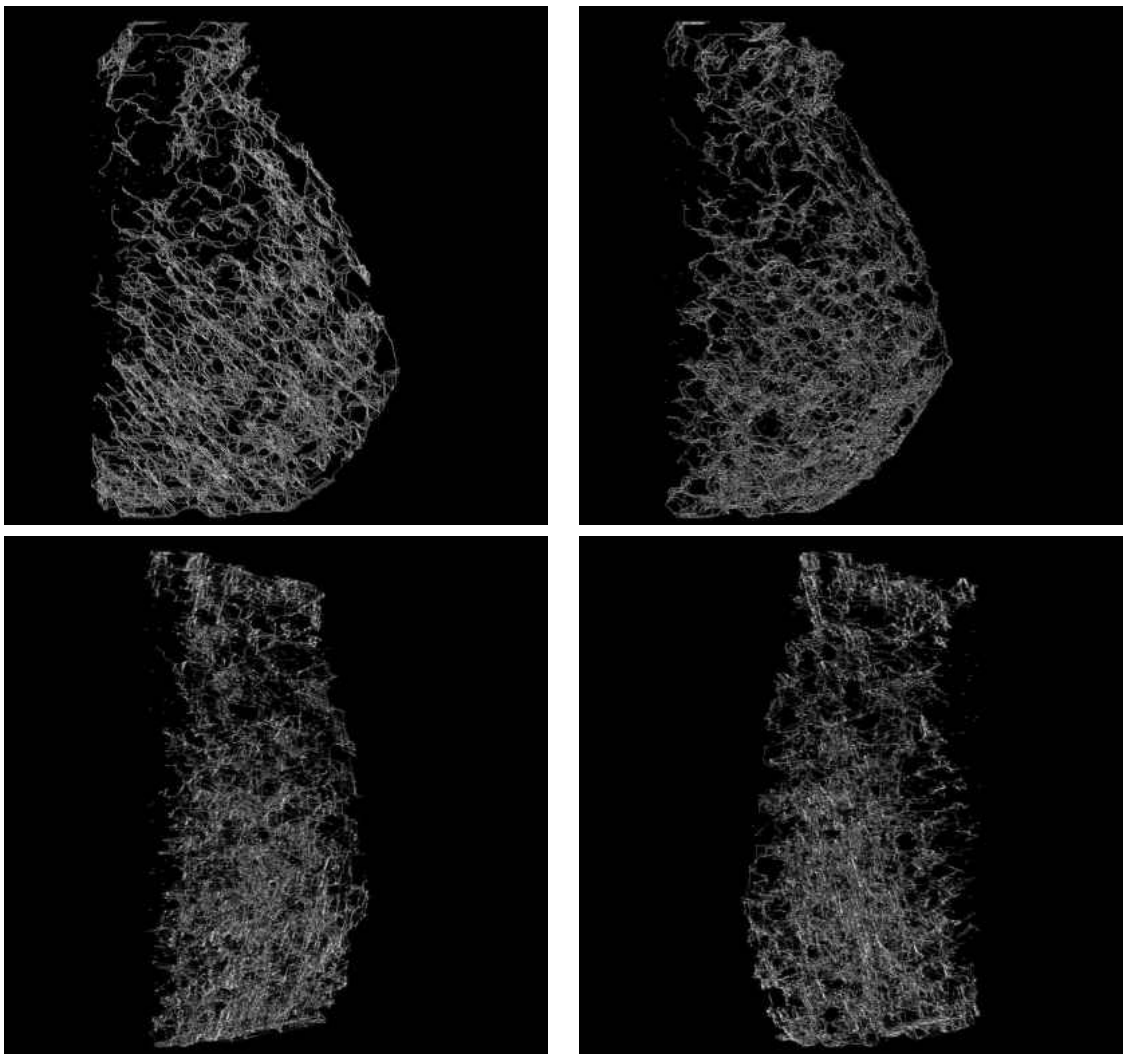


Figure 4.27: Rendering of a typical 3–dimensional network shown at various orientations (continued from Fig. 4.26).

4.5.2 Data

The tomosynthesis image volumes described in Sec. 4.2.2 and Sec. 4.3.2 were used during this classification experiment. This dataset consisted of 80 image volumes comprising left and right tomosynthesis images from 40 women.

The images have been classified according to BIRADS density classes by an expert radiologist. Of the 80 images, there were no images in BIRADS class 1, 21 in BIRADS class 2, 55 in BIRADS class 3 and 4 in BIRADS class 4.

4.5.3 Method

The purpose of this experiment is to identify whether data contained within the 3D network representations produced in Sec. 4.4 can be used in mammographic risk classification to produce an automatic classifier. A non–linear Support Vector Machine (SVM) based classifier was used (see Sec. 3.5.2.4) since this produced the best results during classification of 2D images.

Several features were measured from the network representations produced from each image volume. These attempt to represent not only the density of linear structures present, but also their distribution. They were as follows:

- total number of nodes connected to at least one edge,
- total number of edges,
- mean edge length (distance between connected nodes),
- standard deviation of edge lengths,
- total length of edges.

The feature factors were initially normalised with respect to the breast area. In order to give values for each feature factor that are in similar range so as not to bias a classifier towards particular factors, the values for each factor were then normalised such that the mean value was 0 and the standard deviation was 1.

The normalised values were used in a leave–one–woman–out classification experiment.

4.5.4 Results

The results of the classification are shown in Tab. 4.1. The results show that there is a considerable variation in the performance for different BIRADS classes. In an attempt to explain the observed results, a graph is plotted in Fig. 4.28 showing two feature factors and points grouped by BIRADS class.

Table 4.1: Classification according to BIRADS class using the full dataset.

Placed Class	**True Class**		
	2	**3**	**4**
2	11	3	2
3	10	52	2
4	0	0	0
Proportion correct	0.5238	0.9455	0.0000
		0.7875	
κ **coefficient**	0.68		

The limitations of the dataset become apparent from the scatter–plot. The dataset is very uneven in that approximately one quarter of the images belong to BIRADS class 2, almost three quarters belong to class 3 whilst only 4 images (from two women) belong to class 4 and no images belong to class 1. The graph in Fig. 4.28 shows clearly why the classification for BIRADS class 4 is so poor, since the four images were from two women and each pair of images were a considerable distance from the other pair on the graph. Since the classification experiment was a leave–one–woman–out classification, the remaining pair of images were classified

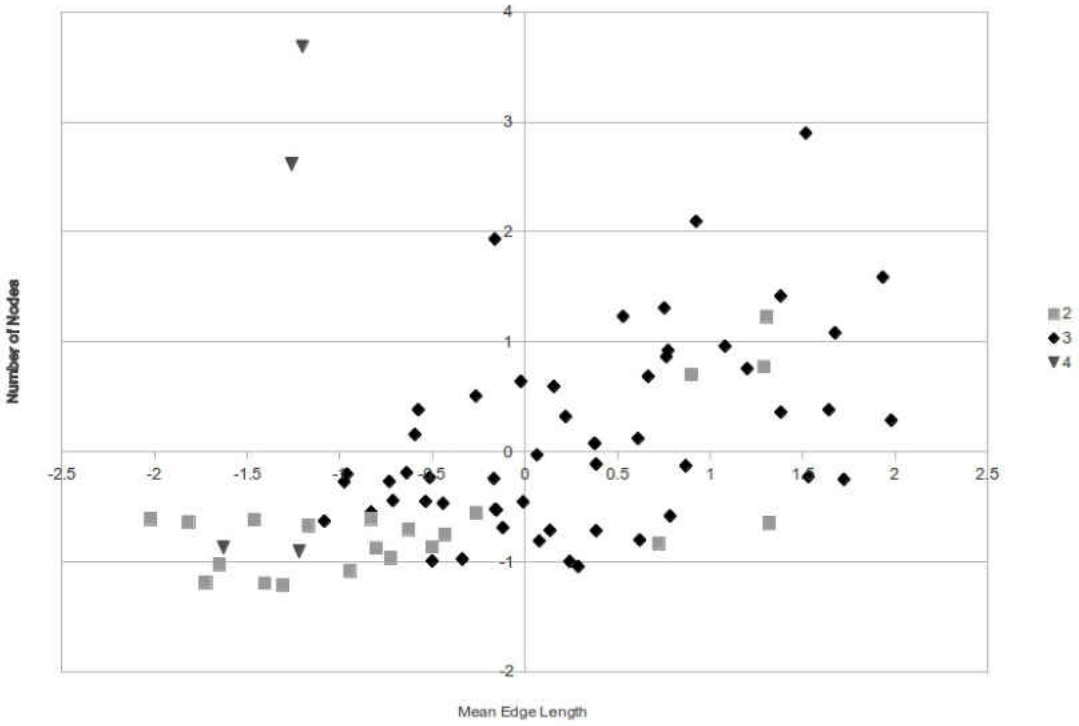


Figure 4.28: A graph comparing two feature factors showing the distribution of BIRADS classes (2, 3 and 4).

as those surrounding them. Similarly, 21 images belonging to class 2 represents a small training set for classification purposes.

In order to remove the effects of the very small set of class 4 images, the images were grouped into high and low risk groups as is common in risk classification. The low risk group comprised BIRADS class 2, whilst the high risk group comprised BIRADS classes 3 and 4. The results of this classification are shown in Tab. 4.2.

Table 4.2: Classification according to high and low risk groups using the full dataset.

Placed Class	**True Class**	
	Low	**High**
Low	12	8
High	9	51
Proportion correct	0.5714	0.8644
	0.7875	
κ **coefficient**	0.58	
Normalised by class proportion	0.7179	

The grouping of images into low and high risk categories gave the same overall classification result, however the results for the individual groups were more closely aligned than with the three BIRADS classes.

Since the left and right images are known to be similar, it was felt that this could have effects on the development of a training set by the classifier. Although the experiment was performed in a leave–one–woman fashion such that both images for each woman were removed from the training set when performing a classification on either image, the remaining data was likely to be largely grouped in pairs such that each data point may be almost duplicated. In order to investigate any effects of this, the dataset was divided into left and right images and a classification experiment was conducted using each half of the dataset. The results of the classification using left images are show in Tabs. 4.3–4.4 and the results of the classification using right images are shown in Tabs. 4.5–4.6.

Table 4.3: Classification according to BIRADS class using left images.

Placed Class	**True Class**		
	2	**3**	**4**
2	6	0	1
3	5	27	1
4	0	0	0
Proportion correct	0.5455	1.0000	0.0000
	0.8250		
κ **coefficient**	0.74		

Table 4.4: Classification according to high and low risk groups using left images.

Placed Class	**True Class**	
	Low	**High**
Low	5	3
High	6	26
Proportion correct	0.4545	0.8966
	0.7750	
κ **coefficient**	0.55	
Normalised by class proportion	0.6755	

The division of the dataset resulted in even smaller sets of images in each category, particularly the BIRADS class 2 / low risk group, resulting in generally poor performance for this group, however the overall classification results are

Table 4.5: Classification according to BIRADS class using right images.

Placed Class	**True Class**		
	2	**3**	**4**
2	4	0	1
3	6	28	1
4	0	0	0
Proportion correct	0.4000	1.0000	0.0000
	0.8000		
κ **coefficient**	0.70		

Table 4.6: Classification according to high and low risk groups using right images.

Placed Class	**True Class**	
	Low	**High**
Low	4	3
High	6	27
Proportion correct	0.4000	0.9000
	0.7750	
κ **coefficient**	0.55	
Normalised by class proportion	0.6500	

comparable to those using the full dataset.

In order to further the investigation into the effects of abnormalities (see Sec. 3.3.4) the image volumes were grouped based on the presence of abnormalities. This resulted in a group of 40 normal images (of which 10 were classified as BIRADS class 2, 28 as BIRADS class 3 and 2 as BIRADS class 4) and 40 abnormal images (of which 11 were classified as BIRADS class 2, 27 as BIRADS class 3 and 2 as BIRADS class 4). The classification results for the normal set of images are shown in Tabs. 4.7–4.8 and the results for the abnormal set are shown in Tabs. 4.9–4.10.

Table 4.7: Classification according to BIRADS class using normal images.

Placed Class	**True Class**		
	2	**3**	**4**
2	6	1	1
3	4	27	1
4	0	0	0
Proportion correct	0.6000	0.9643	0.0000
	0.8250		
κ **coefficient**	0.74		

Table 4.8: Classification according to high and low risk groups using normal images.

Placed Class	**True Class**	
	Low	**High**
Low	5	2
High	5	28
Proportion correct	0.5000	0.9333
	0.8250	
κ **coefficient**	0.65	
Normalised by class proportion	0.7167	

Table 4.9: Classification according to BIRADS class using abnormal images.

Placed Class	**True Class**		
	2	**3**	**4**
2	4	1	1
3	7	26	1
4	0	0	0
Proportion correct	0.3636	0.9630	0.0000
	0.7500		
κ **coefficient**	0.63		

Table 4.10: Classification according to high and low risk groups using abnormal images.

Placed Class	**True Class**	
	Low	**High**
Low	4	2
High	7	27
Proportion correct	0.3636	0.9310
	0.7750	
κ **coefficient**	0.55	
Normalised by class proportion	0.6473	

Results in Tabs. 4.11–4.10 show the linear structure detection and risk classification experiments described in Chapter 3 performed on the set of 80 corresponding 2D mammograms. Results using the full dataset of 80 images are shown in Tab. 4.11. These indicate overall classification accuracies of 71.3% for the 2D mammograms compared with 78.8% for the 3D tomosynthesis image volumes, suggesting that more accurate risk classification may be possible using 3D tomosynthesis images.

Table 4.11: Classification according to BIRADS class using the full dataset of 2D mammograms.

Placed Class	**True Class**		
	2	**3**	**4**
2	10	6	0
3	11	47	4
4	0	2	0
Proportion correct	0.4762	0.8545	0.0000
	0.7125		
κ **coefficient**	0.57		

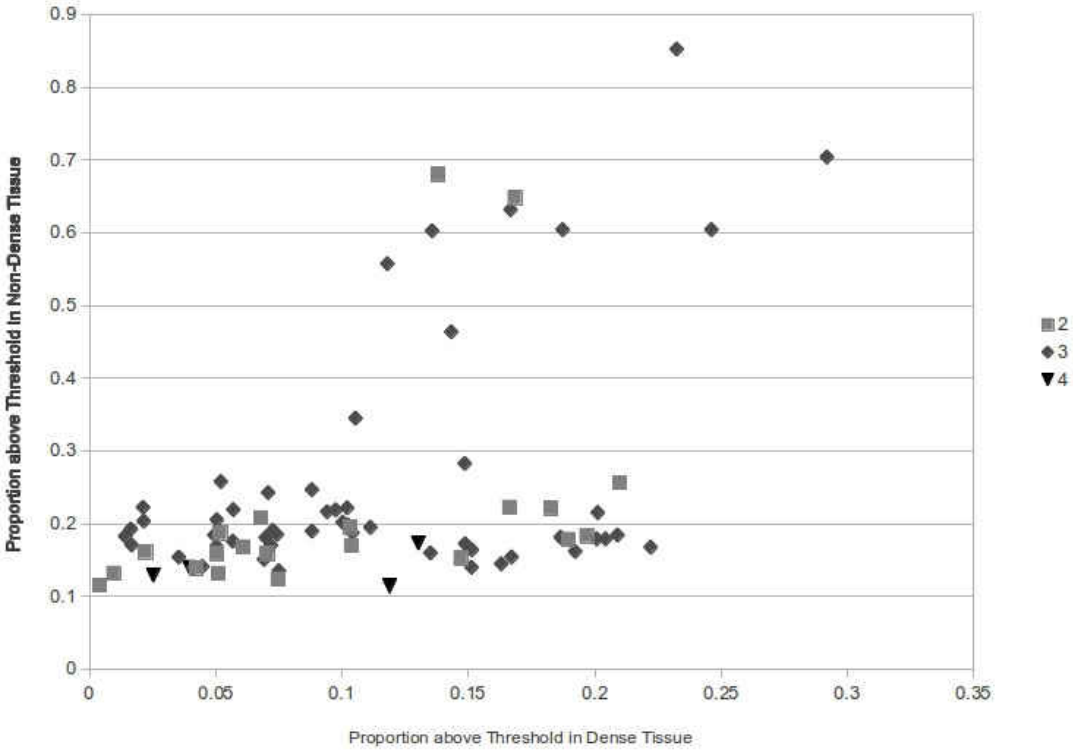


Figure 4.29: A graph comparing two feature factors extracted from 2D mammograms showing the distribution of BIRADS classes (2, 3 and 4).

The graph shown in Fig. 4.29 compares two classification feature factors extracted from each image set. It can be seen that the class 2 data points are more tightly clustered for the 2D images than for the 3D images (see Fig. 4.28), leading to poorer classification accuracy.

Table 4.12 shows the experiment repeated by low/high risk groups. The low risk group comprises BIRADS classes 1 and 2 (although there are no BIRADS class 1 images in this dataset) and the high risk group comprises BIRADS classes 3 and 4. The results show an improved overall classification accuracy for 2D mammograms of 77.5%, although this could at least partially be attributed to a greater imbalance between the results of the two classes.

Table 4.12: Classification according to high and low risk groups using the full dataset of 2D mammograms.

Placed Class	**True Class**	
	Low	**High**
Low	8	5
High	13	54
Proportion correct	0.3810	0.9153
	0.7750	
κ **coefficient**	0.55	
Normalised by class proportion	0.6481	

Tables 4.13–4.14 show the classification experiments performed on the subset of left breast images. In both the BIRADS class and low/high risk group experiments the 3D tomosynthesis images provided improved classification accuracy.

Table 4.13: Classification according to BIRADS class using left 2D images.

Placed Class	**True Class**		
	2	**3**	**4**
2	6	2	0
3	5	24	2
4	0	1	0
Proportion correct	0.5455	0.8889	0.0000
	0.7500		
κ **coefficient**	0.63		

Tables 4.15–4.16 show the classification experiments performed on the subset of right breast images. The classification according to BIRADS class showed

Table 4.14: Classification according to high and low risk groups using left 2D images.

Placed Class	**True Class**	
	Low	**High**
Low	5	4
High	6	25
Proportion correct	0.4545	0.8621
	0.7500	
κ **coefficient**	0.50	
Normalised by class proportion	0.6583	

improved accuracy using the 3D tomosynthesis images, however when combined in to low/high risk groups, the 2D and 3D images achieved the same overall classification accuracy.

Table 4.15: Classification according to BIRADS class using right 2D images.

Placed Class	**True Class**		
	2	**3**	**4**
2	3	2	0
3	7	23	2
4	0	3	0
Proportion correct	0.3000	0.8214	0.0000
	0.6500		
κ **coefficient**	0.48		

Table 4.16: Classification according to high and low risk groups using right 2D images.

Placed Class	**True Class**	
	Low	**High**
Low	3	2
High	7	28
Proportion correct	0.3000	0.9333
	0.7750	
κ **coefficient**	0.55	
Normalised by class proportion	0.6167	

The dataset was divided in to subsets of normal and abnormal images and the experiments were repeated on each subset. Table 4.17 shows the results of the classification according to BIRADS class performed on the subset of normal images. Table 4.18 shows the same experiment classified in to low/high risk groups. In both experiments, the 3D tomosynthesis images provided improved classification accuracy.

Table 4.17: Classification according to BIRADS class using normal 2D images.

Placed Class	**True Class**		
	2	**3**	**4**
2	3	6	0
3	7	21	2
4	0	1	0
Proportion correct	0.3000	0.7500	0.0000
	0.6000		
κ **coefficient**	0.40		

Table 4.18: Classification according to high and low risk groups using normal 2D images.

Placed Class	**True Class**	
	Low	**High**
Low	3	6
High	7	24
Proportion correct	0.3000	0.8000
	0.6750	
κ **coefficient**	0.35	
Normalised by class proportion	0.5500	

Tables 4.19–4.20 show the results of experiments performed on the subset of abnormal images classified according to BIRADS class and low/high risk groups respectively. The 3D tomosynthesis images provided improved classification accuracy in both experiments.

Table 4.19: Classification according to BIRADS class using abnormal 2D images.

Placed Class	**True Class**		
	2	**3**	**4**
2	5	6	0
3	6	19	2
4	0	2	0
Proportion correct	0.4545	0.7037	0.0000
	0.6000		
κ **coefficient**	0.40		

4.5.5 Discussion

The classification results shown in Tab. 4.1 show a high degree of classification accuracy (78.75%) especially given the small dataset of 80 images involved in

Table 4.20: Classification according to high and low risk groups using abnormal 2D images.

Placed Class	**True Class**	
	Low	**High**
Low	5	7
High	6	22
Proportion correct	0.4545	0.7586
	0.6750	
κ **coefficient**	0.35	
Normalised by class proportion	0.6066	

this experiment. This would suggest that features relating to the density and distribution of linear structures can be extracted from digital breast tomosynthesis images and used as a good indicator for mammographic risk. This result would seem to support recent studies showing a strong correlation between parenchymal patterns and breast density in breast tomosynthesis images [60].

Other studies involving automated risk assessment have generally been performed using 2D mammograms and are typically density or texture–based, and so it is not straightforward to make a comparison with these results. The classification based on 2D mammograms presented in Sec. 3.5.2.4 achieved a BIRADS classification accuracy of 74%, however this involved a considerably larger dataset of 320 images.

He *et al.* [113, 114] achieved a BIRADS classification accuracy of 75% when using a texture–based classifier on a set of 320 mammograms, whereas Oliver *et al.* [97] achieved a classification accuracy of 86% using a density–based features in a Bayesian classifier, also on a set of 320 mammograms.

A problem with the results shown in Tab. 4.1 is the imbalance in the classification accuracy of the different classes. The BIRADS class 4 images can be effectively discounted since the dataset contains only four class 4 images from two women. The graph in Fig. 4.28 shows that the class 4 images are grouped in pairs, which correspond to each woman, and that the pairs of images are some distance

from one another. As such, when both images from one woman are removed to create a training set, there is little chance of those images being correctly placed in BIRADS class 4 since the only remaining class 4 images in the training set are a considerable distance away. As such, the results for the high/low risk groups may be more appropriate, since the small number of class 4 images are absorbed in to the high risk group along with the class 3 images. In order to make any satisfactory analysis of the performance within class 4, a significantly larger dataset would be required.

The classification accuracy of BIRADS class 2 images is also poor when compared with the class 3 accuracy. Fig. 4.28 shows that whilst there is a substantial group of the class 2 images in the lower left of the graph, there are also several spread throughout the graph in areas surrounded by class 3 images. These distributed images are likely to be classified as class 3 images. Unfortunately, the available dataset contained only a very small set of class 2 images (21) and so it would be desirable to repeat the experiment using a larger and more balanced dataset before conclusions about the accuracy of the classification of class 2 images can be made.

A clear deficiency of the available dataset was the complete lack of BIRADS class 1 images. This has two principal effects on the results, firstly that we are unable to assess BIRADS class 1 classification, and secondly that this may have improved the performance of the remaining classes since it has effectively removed one the the four possible classification options.

Table 4.2 shows the results of the classification when the images are combined into low (BIRADS 1–2) and high (BIRADS 3–4) risk groups. The overall accuracy remains unchanged although the performance of the two classes is slightly more balanced. Since the only change is that the four BIRADS class 4 images are combined with the BIRADS class 3 images, it was expected that the difference

would be limited. Absorbing the small number of class 4 images in to a larger class allows for a more meaningful normalisation according to relative class size, yielding an overall classification accuracy of 72%.

Since many of the shortcomings of this experiment are due to the limited dataset available, further investigation would ideally involve a repeat of this experiment using a larger and more balanced dataset. Such data is currently difficult to obtain due to the experimental nature of digital breast tomosynthesis imaging [115, 116].

Tables 4.3–4.6 show the results of classification experiments performed on the subsets of left and right images. This was intended to examine any possible effects of including two images from each woman in the training dataset. Although the classification was performed in a leave–one–woman out fashion, whereby both images for a woman are removed when building a training set for placing each image for the given woman, it was suggested that since there is known to be a high degree of similarity between left and right mammographic images for the same woman that this may have a distortive effect on the training set and subsequent classification.

The results shown in Tabs. 4.3–4.6 are similar to the results shown in Tabs. 4.1–4.2 for the full dataset, with any differences unlikely to be significant when considering the very small datasets now involved. This suggests that there are no distortive effects created by including both images for each woman in the dataset.

The final results matrices, shown in Tabs. 4.7–4.10, show the results of the classification performed on the subsets of normal and abnormal images. The investigation in Sec. 3.3.4 concluded that there was no difference in the performance of a classification performed on 2D mammograms when using normal or abnormal images. This experiment intended to investigate whether the same conclusion could be drawn from a classification performed on 3D tomosynthesis images.

The results in Tabs. 4.7 and 4.9 show an improved classification accuracy when using normal images compared with abnormal images (82.5% *vs.* 75%). Whilst the results for BIRADS class 3 images were almost identical, with 27 (96.4%) of 28 normal images correctly classified and 26 (96.3%) of 27 abnormal images correctly classified, the results for BIRADS class 2 images showed a marked difference. The results showed that 6 (60%) of 10 normal BIRADS class 2 images were correctly classified whilst only 4 (36.4%) of 11 abnormal class 2 images were correctly classified.

Unfortunately due to the very small numbers of images involved it is not plausible to draw any firm conclusions from these results, and a repeat of the experiment using a substantially larger dataset would be required before such conclusions could be made. As is shown in Tabs. 4.8 and 4.10 displaying results of the same dataset grouped into low and high risk groups, a change in the classification results of a very small number of images can make an apparently large difference to the classification results. The low risk classification of abnormal images remained that 4 (36.4%) of 11 images were correctly classified, however only 5 (50%) of 10 normal low risk images were correctly classified, suggesting that the observed differences are unlikely to be significant given the small datasets.

The classification results shown in Tab. 4.11 show an overall classification accuracy of 71.3% for 2D mammograms compared with the result of 78.8% for 3D tomosynthesis image volumes shown in Tab. 4.1. Overall, ten classification experiments were performed on each dataset when taking in to account the experiments performed on the full dataset and the subsets of left, right, normal and abnormal images, and for each of these the classification in to BIRADS class and low/high risk groups. In nine of ten experiments the 3D tomosynthesis images provided improved classification accuracy whilst the two datasets performed equally in the remaining one experiment. This would suggest that the method of extracting linear structure information from 3D tomosynthesis images has the potential to

provide more accurate risk assessments than a classifier based on 2D mammograms.

Recent studies have suggested that 3D tomosynthesis images may provide improved sensitivity to breast cancer detection and improved performance in screening to detect possible abnormalities [117, 118, 116], however there has been relatively little work performed investigating risk assessment using tomosynthesis volumes. It had been hypothesised that since tomosynthesis has the potential to provide more information on the parenchymal patterns within the breast that tomosynthesis could provide more accurate automated risk assessment, and the results shown in Tabs. 4.11–4.20 would appear to support this, although further studies using larger datasets would be required before any reliable conclusions could be made.

The graphs shown in Figs. 4.29 and 4.28 provide an overview of the data points for the 2D and 3D datasets respectively. It is clear from Fig. 4.29 that the BIRADS class 2 images are more spread out than those in Fig. 4.28, which are predominantly contained within a well–defined cluster. This could be due to the inherent limitations of 2D mammography such that when a 3D breast is projected on to a 2D image the differences in linear structure density will be less well–defined than when the breast is viewed in a 3D image volume, since all linear structures in the Z–axis will appear overlapping. The effect of this can be seen in the consistently poorer classification accuracy for BIRADS class 2 images in the 2D mammograms compared with the 3D tomosynthesis image volumes.

There is a notable difference between the results based on classification according to BIRADS class and the results based on classification according to low/high risk groups. The experiments using the 2D mammograms generally show improved performance when classified in to low/high risk groups (77.5% for low/high risk groups, 71.3% for BIRADS class using full dataset), whereas the experiments using

3D tomosynthesis images show similar or equal performance (78.8% for both experiments using full dataset). An improvement in performance would be expected since the dataset is classified in to fewer classes, however it would be expected in both datasets. Since the only difference between the results for these two classification approaches is the placement of four images from two women, it is impossible to draw any conclusions from this observation at present, and further study involving a larger and more evenly distributed dataset would be required in order to investigate whether this is a general trend or simply an anomaly in the current dataset.

The experiments performed on the subset of left images showed similar results to those performed on the full dataset. The 2D mammograms provided a classification accuracy of 75.0% (for both BIRADS and low/high risk groups) and the 3D tomosynthesis volumes provided a classification accuracy of 82.5% (77.5% for low/high risk groups). The experiments performed on the subset of right images again showed similar results, with the 2D mammograms providing a classification accuracy of 65.0% (77.5% for low/high risk groups) and the 3D tomosynthesis images providing a classification accuracy of 80.0% (77.5% for low high risk groups). The difference between the BIRADS class and low/high risk group classification experiments is particularly apparent in these results, where there is a large difference between the 2D and 3D datasets when classified according to BIRADS class (65.0% and 80.0% respectively), whereas both datasets performed equally when classified according to low/high risk groups (77.5%).

The experiments performed on subsets of normal and abnormal images were also similar. The slight difference between normal and abnormal subsets in the 3D dataset (82.0% for normal, 75.0% for abnormal) is not present in the 2D dataset (60.0% for both subsets), however the difference present in the 3D dataset is very small considering the size of the subsets, representing a difference of only a few images.

A trend can be seen when examining the classification accuracy of individual BIRADS classes (or low/high risk groups). The BIRADS class 2/low risk group performs consistently poorer when compared with the BIRADS class 3/high risk group, across both 2D and 3D datasets. This could be at least partly due to the uneven nature of the dataset, which contains only 21 BIRADS class 2 (low risk) images compared with 55 BIRADS class 3 and 4 BIRADS class 4 images (59 high risk images in total). Further investigation using a larger and more balanced dataset would be a considerable improvement and would be required in order to draw more substantive conclusions.

Chapter 5

Conclusions

5.1 Evaluation of Aims & Objectives

Section 1.2 referred to six principal research objectives within the project:

1. select a suitable mammographic line detection methodology,
2. investigate the relationship between the density of linear structures in 2D mammograms and mammographic risk,
3. detect mammographic linear structures in raw tomosynthesis images,
4. detect mammographic linear structures in reconstructed tomosynthesis images,
5. develop a model of the distribution of linear structures, and
6. investigate how the distribution of mammographic linear structures might be related to risk.

The first research objective involved the detection of linear structures. This work was important since it forms the foundations for all subsequent experiments,

the success of which depend on the accurate detection of linear structures. Section 3.3 described the implementation of the Line Operator method to detect linear structures in 2D mammographic images, a method that was shown to perform well in its application in subsequent experiments. The only significant challenge faced was that due to lower contrast the linear structure detection was less accurate in dense tissue areas than in non–dense areas, which caused problems for risk classification. This was overcome by the incorporation of dense tissue segmentation and separating linear density into two separate feature factors: linear density in dense tissue and linear density in non–dense tissue. The linear structure detection methodology that was eventually developed was applied successfully in subsequent experiments, suggesting that the first research objective was achieved.

The second research objective involved investigating the relationship between linear density and mammographic risk. Early experiments in Sec. 3.3 demonstrated a significant difference between the linear density in different risk categories of images, suggesting that the relationship between linear density and mammographic risk described by Tabár [20, 21, 22] could be used in automated risk assessment.

Section 3.5 investigated the automatic classification of 2D mammograms into risk categories. The investigation involved the use of two different classifiers: linear discriminant analysis and a non–linear support vector machine–based classifier, and compared results of classification based on linear structure information alone, density information alone and a combination of linear structure and density information. The classification based on linear structure information was shown to out–perform the classification based on density information (64% correct BIRADS classification using linear density compared with 53% using breast density), whilst the combined classification out–performed both individual classifications (74% correct BIRADS classification). This suggests that there exists a relationship between linear structure information and mammographic risk that

can provide useful information for mammographic risk assessment, and therefore that the second research objective has been achieved.

Research objective 3 described an intermediate step prior to investigations using 3D tomosynthesis images. Since 'raw' 2D images used to construct 3D tomosynthesis image volumes are commonly captured using a significantly lower x–ray dose than is used for conventional mammograms, concern was raised that the lower–contrast images may not contain sufficient detail for accurate linear structure detection. An investigation described in Sec. 4.2 applies the technique developed in Chapter 3 for detecting linear structures in 2D mammograms to the 2D 'raw' tomosynthesis images. The experiment used mammograms and 'raw' tomosynthesis images from the same patients acquired during the same screening session and showed that the linear densities measured from the two sets of images were well–correlated (0.866). This suggested that sufficient linear structure information was captured using the lower x–ray dose and therefore that there was likely to be sufficient linear structure information present in 3D tomosynthesis image volumes.

Section 4.3 described the development of a modified Line Operator for use in 3D tomosynthesis image volumes. The detection of linear structures in 3D images proved to be less straightforward than previously expected, and initial results were poor (see Sec. 4.4.2.2). This was largely due to artefacts and distortions in the images introduced in the reconstruction process and required the incorporation of several additional steps into the detection algorithm (see Secs 4.4.2.3–4.4.2.4). The developed approach was used successfully in further experiments, suggesting that research objective 4 has been achieved.

The fifth research objective was to investigate the modelling of the distribution of linear structures. The investigation was described in Sec. 4.4 and involved creating a three–dimensional graph representation of the linear structure tree found

within the 3D tomosynthesis image volumes. From this graph representation, information could be extracted describing the distribution of the linear structures. The developed approach used maximum line strength points as nodes in the graph, since these proved relatively straightforward to detect. These nodes were then connected to nearby nodes using an A*–based search algorithm. During the experiment significant shortcomings were identified in the quality of the line detection results as described in research objective 4 above, largely as a result of artefacts and distortions in the 3D tomosynthesis image volumes. This necessitated the introduction of several additional steps to minimise the effect of these distortions on the linear structure detection results so that a graph representation could be extracted. The development of the graph model of linear structure distribution in preparation for its application in mammographic risk classification suggests that the fifth research objective has been achieved.

The final research objective was to investigate how the distribution of linear structures might be related to mammographic risk. The investigation described in Sec. 4.5 used information from the 3D graph representation described in research objective 5 (see Sec. 4.4) as feature factors for a support vector machine–based classifier. The set of images was classified according to the BIRADS metric and overall results obtained showed a high degree of classification accuracy (79% correct BIRADS classification), suggesting that the incorporation of information relating to the distribution of linear structures might further improve risk assessment. The application of the graph model in the classification according to mammographic risk suggests that the sixth research objective has been achieved, with the conclusion that the experiments suggest a link between the density of linear structures and mammographic risk and that this relationship can be used to provide useful information for automated mammographic risk assessment.

Section 4.5 includes results to compare the classifier based on 2D mammograms (using information about breast density and the density of linear structures) with

the classifier based on 3D tomosynthesis images (using information about the distribution of linear structures). Results showed that the classifier based on 3D tomosynthesis images outperformed the 2D classifier (79% correct BIRADS classification using 3D tomosynthesis images compared with 71% using 2D mammograms), suggesting that the method of extracting linear structure information from 3D tomosynthesis images has the potential to provide more accurate risk assessments than a classifier based on 2D mammograms.

5.2 Discussion & Future Work

Overall conclusions of this study suggest that mammographic risk and linear structure information are related as suggested by Tabár [20, 21, 22]. The investigations described in Chapter 3 demonstrated that the density of linear structures can be used in a risk classifier to place mammographic images into risk categories with a reasonably high degree of precision. The comparison between classification experiments based on breast density segmentation, linear density and a combination of breast density and linear density showed that the classification was improved substantially by the inclusion of linear structure information. This suggests that linear structure information may provide improved risk assessment when compared with conventional density–based classifiers and that the incorporation of linear structure information brings additional risk–related information that was not identified using density segmentation alone.

The classification experiments using information on the distribution of linear structures extracted from 3D tomosynthesis images achieved more accurate classification results, suggesting that the distribution of linear structures is related to mammographic risk and provides additional useful information for risk classification. The experiments based on 2D mammograms showed that a combination of linear structure and density information outperformed each individual factor, so

further work may involve the inclusion of information relating to the volume of dense tissue as a feature factor in the classifier.

The investigations using 3D Tomosynthesis Images involved the development of the graph model to represent linear structures within a breast (see Chapter 4). The graph representation was used to extract information on the distribution of linear structures for use in risk classification, however an avenue that was beyond the scope of this project would be to investigate how the 3D graph of linear structures might be used to assist with the computer–aided detection and classification of abnormalities. Since some abnormalities such as spiculated lesions feature patterns of linear structures, some computer–aided detection systems have used linear structure information to help detect these abnormalities [119, 120]. It is possible that the 3D graph representation may improve the detection of certain abnormalities by providing improved information on the distribution of linear structures. Further work in this area could investigate how the presence of abnormalities affects the distribution of linear structures and whether this information can be used to improve computer–aided detection systems. This work may involve the development of a 3D probability map of the distribution of linear structures to investigate how abnormalities could be detected by their effect on the linear structures.

One of the principal limitations of the classification of 3D tomosynthesis images according to mammographic risk was the shortage of data available for classification. At the time of the experiments, tomosynthesis was an experimental technique and whereas there are several large datasets of 2D mammograms available [104, 23] for research use, no similar datasets of 3D tomosynthesis images existed. This may improve in time as since these experiments were conducted, the first breast tomosynthesis systems have been approved for clinical use in the USA, which may allow for further investigation to be conducted using larger datasets to achieve more reliable conclusions.

Further work in the area of mammographic risk assessment may involve investigating different classifiers. Oliver *et al.* [96] found that an SVM–based classifier did not perform as well as expected during a classification experiment based on texture information from 2D mammograms, and that a classifier based on K–Nearest–Neighbour with Sequential Forward Selection (SFS) feature selection substantially outperformed all others used, including an SVM–based classifier and Linear Discriminant Analysis. As such, it may be possible to improve the observed classification accuracy by investigating similar classifiers.

A worthwhile experiment may be to investigate the combination of 2D and 3D results in a risk classifier. Although this is not a realistic clinical option, since the patient would be subjected to double the normal x–ray dose, it may provide additional information as to the relative performance of the 2D and 3D classifiers and would provide additional validation as to whether 3D tomosynthesis images are able to provide more useful information than 2D mammograms.

An additional improvement would be to develop a more principled approach to determining suitable thresholds. This could be conducted by analysing the grey–level histogram of an image and determining the threshold automatically using a method such as minimum error thresholding [82] or similar. This approach would remove any subjective judgement from the threshold selection process, and may produce more standardised results across different datasets.

The 3D network representation has the potential to be useful in computer–aided diagnosis of breast cancer. Some breast cancers, such as spiculated lesions, may be visible as patterns of linear structures which would be visible in the linear structure network. Some may be difficult to see but may cause architectural distortions [121, 122]. These architectural distortions are difficult to detect, however it is possible that they may affect the distribution of the surrounding linear structures which may be reflected in the 3D network model. As such, further work

may investigate the application of the 3D network model in the detection of breast cancers.

Appendix A

Publications

Work described in Chapter 3 has been previously published in the following papers:

- E. M. Hadley, E. R. E. Denton, and R. Zwiggelaar. Mammographic risk assessment based on anatomical linear structures. *Lecture Notes in Computer Science*, 4046:626–633, 2006 [123].
- E. M. Hadley, E. R. E. Denton, and R. Zwiggelaar. Mammographic risk assessment based on anatomical linear structures: Do mammographic abnormalities play a role? In *Medical Image Understanding and Analysis 2006*, volume 1, pages 11–15, Manchester, UK, 2006 [124].
- E. M. Hadley, E. R. E. Denton, J. Pont, E. Pérez and R. Zwiggelaar. Risk classification of mammograms using anatomical linear structure and density information. *Lecture Notes in Computer Science*, 4078:186–193, 2007 [106].
- E. M. Hadley, E. R. E. Denton, J. Pont, E. Pérez and R. Zwiggelaar. Analysis of Anatomical Linear Structure Information in Mammographic Risk Assessment. *Lecture Notes in Computer Science*, 5116:483–490, 2008 [107].
- E. M. Hadley, E. R. E. Denton, J. Pont, E. Pérez and R. Zwiggelaar. Mammographic Risk Assessment: Does Anatomical Linear Structure Information

Improve the Accuracy of a Density–Based Classifier? In *Medical Image Understanding and Analysis 2008*, pages 224–228, Dundee, UK, 2008 [125].

The first paper [123] was published in Lecture Notes in Computer Science and presented orally at the International Workshop for Digital Mammography (IWDM) 2006 in Manchester on 18^{th}–21^{st} June 2006. This contains the initial linear structure detection work described in Sec. 3.3. The second paper [124] was presented orally at Medical Image Understanding and Analysis (MIUA) 2006 in Manchester on 4^{th}–5^{th} July 2006, and published in the conference proceedings. This second paper contains a more summarised version of the fundamental aspects of the work described in section 3.3, together with a subsequent investigation in to the effect of abnormalities in the data set (see section 3.3).

The third paper [106] was presented orally at the 3^{rd} Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA) 2007 in Girona, Spain on 6^{th}–8^{th} June 2007 and published in Lecture Notes in Computer Science. This paper contains the density estimation work described in section 3.4 and the fundamental linear classification described in section 3.5. Work described in the fourth and fifth papers [107, 125] builds upon this foundation, using more complex non–linear classification methods and a larger dataset (see section 3.5).

The fourth paper [107] was published in Lecture Notes in Computer Science and presented as a poster on 20^{th}–23^{rd} July 2008 at the International Workshop for Digital Mammography (IWDM) 2008 in Tucson, Arizona, USA. The fifth paper [125] was be presented as a poster at Medical Image Understanding and Analysis (MIUA) 2008 in Dundee on 2^{nd}–3^{rd} July 2008 and published in the conference proceedings.

Work described in Chapter 4 has been published in the following peer–reviewed papers:

- E. M. Hadley, P. R. Bakic, A. D. A. Maidment and R. Zwiggelaar. Detection of anatomical linear structures in mammograms and raw tomosynthesis images. In *Medical Image Understanding and Analysis 2007*, 71–75, 2007 [126].

- E. M. Hadley, P. R. Bakic, A. D. A. Maidment and R. Zwiggelaar. Detection of linear structures in 3D breast tomosynthesis images. In *Medical Image Understanding and Analysis 2009*, 194–198, 2009 [127].

The first paper describes initial work involving correspondence between conventional mammograms and 'raw' projection (unreconstructed) tomosynthesis images. This was presented as a poster at Medical Image Understanding and Analysis (MIUA) 2007 in Aberystwyth on 17^{th}–18^{th} July 2007 and published in the conference proceedings. The second paper describes the approach taken to detect linear structures in 3D reconstructed tomosynthesis images, an essential step for future work in this area. This paper was presented as a poster at MIUA 2009 in Kingston, London on 14^{th}–15^{th} July 2009 and published in the conference proceedings.

Bibliography

[1] Ahmedin Jemal, Freddie Bray, Melissa M. Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61(2):69–90, 2011.

[2] D. Max Parkin, Freddie Bray, J. Ferlay, and Paola Pisani. Global cancer statistics, 2002. *CA: A Cancer Journal for Clinicians*, 55(2):74–108, 2005.

[3] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M. J. Thun. Cancer statistics, 2008. *CA: A Cancer Journal for Clinicians*, 58(2):71–96, 2008.

[4] Rebecca Siegel, Elizabeth Ward, Otis Brawley, and Ahmedin Jemal. Cancer statistics, 2011. *CA: A Cancer Journal for Clinicians*, 61(4):212–236, 2011.

[5] SH Landis, T Murray, S Bolden, and PA Wingo. Cancer statistics 1998. *CA: A Cancer Journal for Clinicians*, 48:6–31, 1998.

[6] J Ferlay, P Autier, M Boniol, M Heanue, M Colombet, and P Boyle. Estimates of the cancer incidence and mortality in europe in 2006. *Annals of Oncology*, 18(3):581–592, 2007.

[7] László Tabár, Bedrich Vitak, Tony Hsiu-Hsi Chen, Amy Ming-Fang Yen, Anders Cohen, Tibor Tot, Sherry Yueh-Hsia Chiu, Sam Li-Sheng Chen, Jean Ching-Yuan Fann, Johan Rosell, Helena Fohlin, Robert A. Smith, and Stephen W. Duffy. Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*, 260:658–663, 2011.

[8] Eitan Amir, Orit C. Freedman, Bostjan Seruga, and D. Gareth Evans. Assessing women at high risk of breast cancer: A review of risk assessment models. *Journal of the National Cancer Institute*, 102(10):680–691, 2010.

[9] Mitchell H. Gail, Louise A. Brinton, David P. Byar, Donald K. Corle, Sylvan B. Green, Catherine Schairer, and John J. Mulvihill. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, 81(24):1879–1886, 1989.

[10] JP Costantino, MH Gail, D Pee, S Anderson, CK Redmond, J Benichou, and HS Wieand. Validation studies for models projecting the risk of invasive

and total breast cancer incidence. *Journal of the National Cancer Institute*, 91:1541–1548, 1999.

[11] Stephen R. Aylward, Bradley M. Hemminger, and Etta D. Pisano. Mixture modeling for digital mammogram display and analysis. *Digital Mammography*, 13:305–312, 1998.

[12] Etta D. Pisano, Constantine Gatsonis, Edward Hendrick, Martin Yaffe, Janet K. Baum, Suddhasatta Acharyya, Emily F. Conant, Laurie L. Fajardo, Lawrence Bassett, Carl D'Orsi, Roberta Jong, and Murray Rebner. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, 353(17):1773–1783, 2005.

[13] T. M. Kolb, J. Lichy, and J. H. Newhouse. Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: An analysis of 27,825 patient evaluations. *Radiology*, 225(1):165–175, 2002.

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

[15] J. Brisson, F. Merletti, N. L. Sadowski, J. A. Twaddle, A. S. Morrison, and P. Cole. Mammographic features of the breast and breast cancer risk. *American Journal of Epidemiology*, 115:428–437, 1982.

[16] J. W. Byng, M. J. Yaffe, R. A. Jong, R. S. Shumak, G. A. Lockwood, D. L. Tritchler, and N. F. Boyd. Analysis of mammographic density and breast cancer risk from digitised mammograms. *Radiographics*, 18(6):1587–1598, 1998.

[17] C. H. van Gils, J. D. Otten, J. H. Hendriks, R. Holland, H. Straatman, and A. L. Verbeek. High mammographic breast density and its implications for the early detection of breast cancer. *Journal of Medical Screening*, 6:200–204, 1999.

[18] M. J. Yaffe, N. F. Boyd, J. W. Byng, R. A. Jong, E. Fishell, G. A. Lockwood, L. E. Little, and D. L. Tritchler. Breast cancer risk and measured mammographic density. *European Journal of Cancer Prevention*, 7(suppl 1):S47–S55, 1998.

[19] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe. Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine*, 356(3):227–236, 2007.

[20] L. Tabár and P. B. Dean. Mammographic parenchymal patterns. risk indicator for breast cancer? *Journal of the American Medical Association*, 247(2):185–189, 1982.

[21] I. T. Gram, E. Funkhouser, and L. Tabár. The Tabár classification of mammographic parenchymal patterns. *European Journal of Radiology*, 24(2):131–136, 1997.

[22] L. Tabár, T. Tot, and P. B. Dean. *Breast Cancer - The Art and Science of Early Detection with Mammography.* Georg Thieme Verlag, Stuttgart, 2005.

[23] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage. The mammographic images analysis society digital mammogram database. *Exerpta Medica. International Congress Series.*, 1069:375–378, 1994.

[24] L. T. Nikalason, B. T. Christian, L. E. Nikalason, D. B. Kopans, D. E. Castleberry, B. H. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore, D. Albagli, M. C. DeJule, P. F. Fitzgerald, D. F. Fobare, B. W. Giambattista, R. F. Kwasnick, J. Liu, S. J. Lubowski, G. E. Possin, J. F. Richotte, C. Y. Wei, and R. F. Wirth. Digital tomosynthesis in breast imaging. *Radiology*, 205:399–406, 1997.

[25] J. T. Dobbins and D. J. Godfrey. Digital x–ray tomosynthesis: current state of the art and clinical potential. *Physics in Medicine and Biology*, 48(19):R65–106, 2003.

[26] M. Varjonen. Three–dimensional digital breast tomosynthesis in the early diagnosis and detection of breast cancer. *Lecture Notes in Computer Science*, 4046:152–159, 2006.

[27] American College of Radiology. *Illustrated Breast Imaging Reporting and Data System.* American College of Radiology, 3rd edition, 1998.

[28] N. F. Boyd, J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B. Miller, G. A. Lockwood, D. L. Tritchler, and M. J. Yaffe. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *Journal of the National Cancer Institute*, 87:670–675, 1995.

[29] I. Muhimmah, A. Oliver, E. R. E. Denton, J. Pont, E. Perez, and R. Zwiggelaar. Comparison between Wolfe, Boyd, BI-RADS and Tabar based mammographic risk assessment. *Lecture Notes in Computer Science*, 4046:407–415, 2006.

[30] J. A. Harvey and V. E. Bovbjerg. Quantitative assessment of mammographic breast density: Relationship with breast cancer risk. *Radiology*, 230:29–41, 2004.

[31] D. Scutt, G. A. Lancaster, and J. T. Manning. Breast asymmetry and a predisposition to breast cancer. *Breast Cancer Research*, 8(2):R14, 2006.

[32] Valerie A. McCormack and Isabel dos Santos Silva. Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention*, 15(6):1159–1169, 2006.

[33] Celia Byrne, Catherine Schairer, John Wolfe, Navin Parekh, Martine Salane, Louise A. Brinton, Robert Hoover, and Robert Haile. Mammographic features and breast cancer risk: Effects with time, age, and menopause status. *Journal of the National Cancer Institute*, 87(21):1622–1629, 1995.

[34] PB Lam, PM Vacek, BM Geller, and HB Muss. The association of increased weight, body mass index, and tissue density with risk of breast carcinoma in vermont. *Cancer*, 89:369–375, 2000.

[35] NF Boyd, GA Lockwood, JW Byng, DL Tritchler, and MJ Yaffe. Mammographic densities and breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*, 7:1133–1144, 1998.

[36] RL Egan and RC Mosteller. Breast cancer mammography patterns. *Cancer*, 40:2087–2090, 1977.

[37] DS Buist, PL Porter, C Lehman, SH Taplin, and E White. Factors contributing to mammography failure in women aged 40-49 years. *Journal of the National Cancer Institute*, 96:1432–1440, 2004.

[38] Sherry Yueh-Hsia Chiu, Stephen Duffy, Amy Ming-Fang Yen, Laszlo Tabr, Robert A. Smith, and Hsiu-Hsi Chen. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a swedish mammographic screening. *Cancer Epidemiology, Biomarkers & Prevention*, 19(5):1219–1228, 2010.

[39] L Tabár, CJ Fagerberg, and A Gad. Reduction in mortality from breast cancer after mass screening with mammography. randomised trial from the breast cancer screening working group of the swedish national board of health and welfare. *Lancet*, 1:829–832, 1985.

[40] Edward A. Sickles. Wolfe mammographic parenchymal patterns and breast cancer risk. *American Journal of Roentgenology*, 188(2):301–303, 2007.

[41] Therese B. Bevers, Deborah K. Armstrong, Banu Arun, Robert W. Carlson, Kenneth H. Cowan, Mary B. Daly, Irvin Fleming, Judy E. Garber, Mary Gemignani, William J. Gradishar, Helen Krontiras, Swati Kulkarni, Christine Laronga, Loretta Loftus, Deborah J. MacDonald, Martin C. Mahoney, Sofia D. Merajver, Ingrid Meszoely, Lisa Newman, Elizabeth Pritchard, Victoria Seewaldt, Rena V. Sellin, Charles L. Shapiro, and John H. Ward. Breast cancer risk reduction. *Journal of the National Comprehensive Cancer Network*, 8(10):1112–1146, 2010.

[42] Adriano Decarli, Stefano Calza, Giovanna Masala, Claudia Specchia, Domenico Palli, and Mitchell H. Gail. Gail model for prediction of absolute risk of invasive breast cancer: Independent evaluation in the florenceeuropean prospective investigation into cancer and nutrition cohort. *Journal of the National Cancer Institute*, 98(23):1686–1693, 2006.

[43] William E. Barlow, Emily White, Rachel Ballard-Barbash, Pamela M. Vacek, Linda Titus-Ernstoff, Patricia A. Carney, Jeffrey A. Tice, Diana S. M. Buist, Berta M. Geller, Robert Rosenberg, Bonnie C. Yankaskas, and Karla Kerlikowske. Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute*, 98(17):1204–1214, 2006.

[44] Melanie R. Palomares, Joelle R.B. Machia, Constance D. Lehman, Janet R. Daling, and Anne McTiernan. Mammographic density correlation with gail model breast cancer risk estimates and component risk factors. *Cancer Epidemiology, Biomarkers & Prevention*, 15(7):1324–1330, 2006.

[45] E Amir, D G Evans, A Shenton, F Lalloo, A Moran, C Boggis, M Wilson, and A Howell. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *Journal of Medical Genetics*, 40(11):807–814, 2003.

[46] EB Claus, N Risch, and WD Thompson. Autosomal dominant inheritance of early onset breast cancer: implications for risk prediction. *Cancer*, 73:643–651, 1994.

[47] D Ford, DF Easton, DT Bishop, SA Narod, and DE Goldgar. Risk of cancer in brca-1 mutation carriers. *Lancet*, 343:692–695, 1994.

[48] J Tyrer, SW Duffy, and J Cuzick. A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in Medicine*, 15(7):1111–1130, 2004.

[49] Jeffrey Tice, Steven Cummings, Elad Ziv, and Karla Kerlikowske. Mammographic breast density and the gail model for breast cancer risk prediction in a screening population. *Breast Cancer Research and Treatment*, 94(2):115–122, 2005.

[50] Steven R. Cummings, Jeffrey A. Tice, Scott Bauer, Warren S. Browner, Jack Cuzick, Elad Ziv, Victor Vogel, John Shepherd, Celine Vachon Rebecca Smith-Bindman, and Karla Kerlikowske. Prevention of breast cancer in postmenopausal women: Approaches to estimating and reducing risk. *Journal of the National Cancer Institute*, 101(6):384–398, 2009.

[51] MB Terry, FF Zhang, G Kabat, JA Britton, SL Teitelbaum, AI Neugut, and Gammon MD. Lifetime alcohol intake and breast cancer risk. *Annals of Epidemiology*, 16(3):230–240, 2006.

[52] Bertrand Tehard, Christine M Friedenreich, Jean-Michel Oppert, and Francoise Clavel-Chapelon. Effect of physical activity on women at increased risk of breast cancer: Results from the e3n cohort study. *Cancer Epidemiology, Biomarkers & Prevention*, 15(1):57–64, 2006.

[53] Leslie Bernstein, Alpa V. Patel, Giske Ursin, Jane Sullivan-Halley, Michael F. Press, Dennis Deapen, Jesse A. Berlin, Janet R. Daling, Jill A. McDonald, Sandra A. Norman, Kathleen E. Malone, Brian L. Strom, Jonathan Liff, Suzanne G. Folger, Michael S. Simon, Ronald T. Burkman, Polly A. Marchbanks, Linda K. Weiss, and Robert Spirtas. Lifetime recreational exercise activity and breast cancer risk among black women and white women. *Journal of the National Cancer Institute*, 97(22):1671–1679, 2005.

[54] Prentice RL, Caan B, Chlebowski RT, Patterson R, Kuller LH, Ockene JK, Margolis KL, Limacher MC, Manson JE, Parker LM, Paskett E, Phillips L, Robbins J, Rossouw JE, Sarto GE, Shikany JM, Stefanick ML, Thomson CA, Van Horn L, Vitolins MZ, Wactawski-Wende J, Wallace RB, Wassertheil-Smoller S, Whitlock E, Yano K, Adams-Campbell L, Anderson GL, Assaf AR, Beresford SA, Black HR, Brunner RL, Brzyski RG, Ford L, Gass M, Hays J, Heber D, Heiss G, Hendrix SL, Hsia J, Hubbell FA, Jackson RD, Johnson KC, Kotchen JM, LaCroix AZ, Lane DS, Langer RD, Lasser NL, and Henderson MM. Low-fat dietary pattern and risk of invasive breast cancer: the women's health initiative randomized controlled dietary modification trial. *Journal of the American Medical Association*, 295(6):629–642, 2006.

[55] AH Eliassen, GA Colditz, B Rosner, WC Willett, and SE Hankinson. Adult weight change and risk of postmenopausal breast cancer. *Journal of the American Medical Association*, 296(2):193–201, 2006.

[56] P. T. Huynh, A. M. Jarolimek, and S. Daye. The false-negative mammogram. *Radiographics*, 18(5):1137–1154, 1998.

[57] J. M. Park, E. A. Franken, M. Garg, L. L. Fajardo, and L. T. Nikalason. Breast tomosynthesis: Present considerations and future applications. *Radiographics*, 27:S231–S240, 2007.

[58] T. Wu, R. H. Moore, E. A. Rafferty, and D. B. Kopans. A comparison of reconstruction algorithms for breast tomosynthesis. *Medical Physics*, 31(9):2636–2647, 2004.

[59] A. Smith. Full-field breast tomosynthesis. *Radiology Management*, 27(5):25–31, 2005.

[60] Despina Kontos, Lynda C. Ikejimba, Predrag R. Bakic, Andrea B. Troxel, Emily F. Conant, and Andrew D. A. Maidment. Analysis of parenchymal texture with digital breast tomosynthesis: Comparison with digital mammography and implications for cancer risk assessment. *Radiology*, 261:80–91, October 2011.

[61] Zhimin Huo, Maryellen L. Giger, Carl J. Vyborny, and Charles E. Metz. Breast cancer: Effectiveness of computer-aided diagnosis–observer study with independent database of mammograms. *Radiology*, 224:560–568, 2002.

[62] R. Zwiggelaar, S. M. Astley, C. R. M. Boggis, and C. J. Taylor. Linear structures in mammographic images: Detection and classification. *IEEE Transactions on Medical Imaging*, 23(9):1077–1086, 2004.

[63] R. N. Dixon and C. J. Taylor. Automated asbestos fibre counting. *Institute of Physics Conference Series*, 44:178–185, 1979.

[64] T.C. Parr, C.J. Taylor, S.M. Astley, and C.R.M. Boggis. Statistical modeling of oriented line patterns in mammograms. In *SPIE Medical Imaging*, volume 3034, pages 44–55, 1997.

[65] R. Zwiggelaar, T. C. Parr, and C. J. Taylor. Finding orientated line patterns in digital mammographic images. In *7th Britich Machine Vision Conference*, pages 715–724, 1996.

[66] N. Karssemeijer and G. M. te Brake. Detection of stellate distortions in mammograms. *IEEE Transactions on Medical Imaging*, 15:611–619, 1996.

[67] J. J. Koenderink and A. J. van Doorn. Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:597–605, 1992.

[68] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pat*, 13:891–906, 1991.

[69] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 465–470, San Francisco, CA, USA, June 1996.

[70] T. Lindeberg and M.X. Li. Segmentation and classification of edges using mimimum description length approximation and complementary junction cues. *Computational Vision and Image Understanding*, 67(1):88–98, 1997.

[71] P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment. Mammogram synthesis using a 3d simulation. i. breast tissue model and image acquisition simulation. *Medical Physics*, 29(9):2131–2139, 2002.

[72] P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment. Mammogram synthesis using a 3d simulation. ii. evaluation of synthetic mammogram texture. *Medical Physics*, 29(9):2140–2151, 2002.

[73] P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment. Mammogram synthesis using a three-dimensional simulation. iii. modeling and evaluation of the breast ductal network. *Medical Physics*, 30(7):1914–1925, 2003.

[74] Karla Kerlikowske and Deborah Grady. Variability and accuracy in mammographic interpretation using the american college of radiology breast imaging reporting and data system. *Journal of the National Cancer Institute*, 90(23):1801–1809, 1998.

[75] R. Zwiggelaar, R. Martí, and C. Boggis. The benefit of knowing your linear structures in mammographic images. In *Medical Image Understanding and Analysis '02*, Portsmouth, UK, 2002.

[76] R. Zwiggelaar and C. R. M. Boggis. Classification of linear structures in mammographic images. In *Medical Image Understanding and Analysis '01*, Birmingham, UK, 2001.

[77] R. Zwiggelaar and C. R. M. Boggis. Automatic classification of linear structures in mammographic images - is it possible? In *3rd Conference on Medical Image Understanding and Analysis*, pages 185–188, Oxford, UK, 1999.

[78] N. Cerneaz and M Brady. Finding curvilinear structures in mammograms. *Lectur*, 905:372–382, 1995.

[79] N. F. Boyd, B. O'Sullivan, J. E. Campbell, E. Fishell, I. Simor, G. Cooke, and T. Germanson. Mammographic signs as risk factors for breast cancer. *British Journal of Cancer*, 45(2):185–193, 1982.

[80] John A. Shepherd, Karla Kerlikowske, Lin Ma, Frederick Duewer, Bo Fan, Jeff Wang, Serghei Malkov, Eric Vittinghoff, and Steven R. Cummings. Volume of mammographic density and risk of breast cancer. *Cancer Epidemiology, Biomarkers & Prevention*, 20(7):1473–1482, 2011.

[81] R. Sivaramakrishna, N. A. Obuchowski, W. A. Chilcote, and K. A. Powell. Automatic segmentation of mammographic density. *Academic Radiology*, 8(3):250–256, 1998.

[82] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, 19(1):41–47, 1986.

[83] Chuan Zhou, Heang-Ping Chan, Nicholas Petrick, Berkman Sahiner, Mark A. Helvie, Marilyn A. Roubidoux, Lubomir M. Hadjiiski, and Mitchell M. Goodsitt. Computerized image analysis: estimation of breast density on mammograms. In *SPIE*, volume 3979, page 1615, 2000.

[84] Catherine Klifa, Julio Carballido-Gamio, Lisa Wilmes, Anne Laprie, John Shepherd, Jessica Gibbs, Bo Fan, Susan Noworolski, and Nola Hylton. Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort. *Magnetic Resonance Imaging*, 28(1):8–15, 2010.

[85] Martin Yaffe. Mammographic density. measurement of mammographic density. *Breast Cancer Research*, 10(3):209, 2008.

[86] Yuan-Hsiang Chang, Xiao-Hui Wang, Lara A. Hardesty, Thomas S. Chang, William R. Poller, Walter F. Good, and David Gur. Computerized assessment of tissue composition on digitized mammograms. *Academic Radiology*, 9(8):899–905, 2002.

[87] Carri K. Glide-Hurst, Neb Duric, and Peter Littrup. A new method for quantitative analysis of mammographic density. *Medical Physics*, 34(11):4491, 2007.

[88] Martin KE, Helvie MA, Zhou C, Roubidoux MA, Bailey JE, Paramagul C, Blane CE, Klein KA, Sonnad SS, and Chan H-P. Mammographic density measured with quantitative computer-aided method: Comparison with radiologists' estimates and bi-rads categories. *Radiology*, 240:656–665, 2006.

[89] Yang Can Gong, Michael Brady, and Styliani Petroudi. Texture based mammogram classification and segmentation. *Lecture Notes in Computer Science*, 4046:616–625, 2006.

[90] Jose L. Marroquin, Edgar Arce Santana, and Salvador Botello. Hidden markov measure field models for image segmentation. *IEEE Pattern Analysis and Machine Intelligence*, 25(11):1380–1387, 2003.

[91] S. Petroudi, T. Kadir, and M. Brady. Automatic classification of mammographic parenchymal patterns: a statistical approach. In *IEEE Int. Conf. Engineering in Medicine and Biology Society*, volume 1, pages 798–801, 2003.

[92] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *European Conference on Computer Vision*, pages 255–271, 2002.

[93] K. Bovis and S. Singh. Classification of mammographic breast density using a combined classifier paradigm. In *Proceedings of Medical Image Understanding and Analysis*, pages 177–180, 2002.

[94] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems Man and Cybernetics*, 3(6):610–621, 1973.

[95] M. Heath, K. Bowyer, D. Kopans, W. P. Kegelmeyer, R. Moore, K. Chang, and S. MunishKumaran. Current status of the digital database for screening mammography. In *Proceedings of the Fourth International Workshop on Digital Mammography*, pages 457–460. Kluwer Academic Publishers, 1998.

[96] Arnau Oliver, Xavier Lladó, Robert Martí, Jordi Freixenet, and Reyer Zwiggelaar. Classifying mammograms using texture information. In *Medical Image Understanding and Analysis*, pages 223–227, Aberystwyth, UK, July 2007.

[97] A. Oliver, J. Freixenet, R. Martí, J. Pont, E Pérez, E. R. E. Denton, and R. Zwiggelaar. A novel breast tissue density classification methodology. *IEEE Transactions on Information Technology in BioMedicine*, 12(1):55–65, 2008.

[98] T. Ojala, M. Pietikäinen, and T. Mäenpä. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[99] J. Cohen. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20:27–46, 1960.

[100] J. Landis and G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(3):159–174, 1977.

[101] A. Oliver, J. Freixenet, R. Martí, and R. Zwiggelaar. A comparison of breast tissue classification techniques. *Lecture Notes in Computer Science*, 4191:872–879, 2006.

[102] N. Karssemeijer. Automated classification of mammographic parenchymal pattern. *Physics in Medicine and Biology*, 28:365–378, 1998.

[103] V. N. Vapnik. *The Nature of Staistical Learning Theory*. Springer-Verlag, New York, 1995.

[104] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer. The digital database for screening mammography. In M. J. Yaffe, editor, *Proc. I.W.D.M.*, pages 212–218. Medical Physics Publishing, 2001.

[105] S. Field. Breast screening issues. *Newsletter of the Royal College of Radiologists*, 54:12–14, 1998.

[106] E. M. Hadley, E. R. E. Denton, J. Pont, E. Pérez, and R Zwiggelaar. Risk classification of mammograms using anatomical linear structure and density information. *Lecture Notes in Computer Science*, 4478:186–193, 2007.

[107] E. M. Hadley, E. R. E. Denton, J. Pont, E. Pérez, and R Zwiggelaar. Analysis of anatomical linear structure information in mammographic risk assessment. *Lecture Notes in Computer Science*, 5116:483–490, 2008.

[108] D. Kontos, P. R. Bakic, A. B. Troxel, E. F. Conant, and A. D. A. Maidment. Digital breast tomosynthesis parenchymal texture analysis for breast cancer risk estimation: A preliminary study. *Lecture Notes in Computer Science*, 5116:681–688, 2008.

[109] Yiheng Zhang, Heang-Ping Chan, Berkman Sahiner, Jun Wei, Jun Ge, Lubomir M. Hadjiiski, and Chuan Zhou. Investigation of the z-axis resolution of breast tomosynthesis mammography systems. In *Proc. SPIE 6510, 65104A*, 2007.

[110] Yiheng Zhang, Heang-Ping Chan, Mitchell M. Goodsitt, Andrea Schmitz, Jeffrey W. Eberhard, and Bernhard E. H. Claus. Investigation of different pv distributions in digital breast tomosynthesis (dbt) mammography. *Lecture Notes in Computer Science*, 5116:593–600, 2008.

[111] Kálmán Palágyi and Attila Kuba. A parallel 12-subiteration 3d thinning algorithm to extract medial lines. *Lecture Notes in Computer Science*, 1296:400–407, 1997.

[112] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.

[113] Wenda He, Erika R. E. Denton, and Reyer Zwiggelaar. Mammographic image segmentation and risk classification using a novel texture signature based methodology. *Lecture Notes in Computer Science*, 6136:526–533, 2010.

[114] Wenda He, Erika R. E. Denton, Kirsten Stafford, and Reyer Zwiggelaar. Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments. *Biomedical Signal Processing and Control*, 6(3):321–329, July 2011.

[115] Daniela Bernardi, Stefano Ciatto, Marco Pellegrini, Paolina Tuttobene, Carmine Fanto, Marvi Valentini, Stefano Michele, Paolo Peterlongo, and Nehmat Houssami. Prospective study of breast tomosynthesis as a triage to assessment in screening. *Breast Cancer Research and Treatment*, pages 1–5, 2012.

[116] A. Tingberg, D. Förnvik, S. Mattsson, T. Svahn, R. Timberg, and S. Zackrisson. Breast cancer screening with tomosynthesis–initial experiences. *Radiation Protection Dosimetry*, 147(1–2):180–183, 2011.

[117] Ingvar Andersson, Debra M. Ikeda, Sophia Zackrisson, Mark Ruschin, Tony Svahn, Pontus Timberg, and Anders Tingberg. Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and birads classification in a population of cancers with subtle mammographic findings. *European Radiology*, 18(12):2817–2825, 2008.

[118] Anders Tingberg, Ingvar Andersson, Debra Ikeda, Mark Ruschin, Tony Svahn, and Pontus Timberg. Birads classification in breast tomosynthesis compared to mammography and ultrasonography. *Lecture Notes in Computer Science*, 5116:67–73, 2008.

[119] Kerstin Moberg, Nils Bjurstam, Brigitte Wilczek, Lars Rostgrd, Elisabeth Egge, and Catharina Muren. Computed assisted detection of interval breast cancers. *European Journal of Radiology*, 39(2):104 – 110, 2001.

[120] Jay A. Baker, Eric L. Rosen, Joseph Y. Lo, Edgardo I. Gimenez, Ruth Walsh, and Mary Scott Soo. Computer-aided detection (cad) in screening mammography: Sensitivity of commercial cad systems for detecting architectural distortion. *American Journal of Roentgenology*, 181(4):1083–1088, 2003.

[121] Sickles. Mammographic features of 300 consecutive nonpalpable breast cancers. *American Journal of Roentgenology*, 146:661–663, 1986.

[122] Knutzen AM and Gisvold JJ. Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. *Mayo Clin Proc*, 68:454–460, 1993.

[123] E. M. Hadley, E. R. E. Denton, and R. Zwiggelaar. Mammographic risk assessment based on anatomical linear structures. *Lecture Notes in Computer Science*, 4046:626–633, 2006.

[124] E. M. Hadley, E. R. E. Denton, and R. Zwiggelaar. Mammographic risk assessment based on anatomical linear structures: Do mammographic abnormalities play a role? In *Medical Image Understanding and Analysis 2006*, volume 1, pages 11–15, Manchester, UK, 2006.

[125] E. M. Hadley, E. R. E. Denton, J. Pont, E. Pérez, and R Zwiggelaar. Mammographic risk assessment: Does anatomical linear structure information improve the accuracy of a density–based classifier? In *Medical Image Understanding and Analysis 2008*, pages 224–228, Dundee, UK, 2008.

[126] E. M. Hadley, P. R. Bakic, A. D. A. Maidment, and R. Zwiggelaar. Detection of anatomical linear structures in mammograms and raw tomosynthesis images. In *Medical Image Understanding and Analysis 2007*, pages 71–75, Aberystwyth, UK, 2007.

[127] E. M. Hadley, P. R. Bakic, A. D. A. Maidment, and R Zwiggelaar. Detection of linear structures in 3d breast tomosynthesis images. In *Medical Image Understanding and Analysis 2009*, pages 194–198, Kingston-upon-Thames, UK, 2009.