PATTERN RECOGNITION APPLIED TO THE COMPUTER-AIDED DETECTION AND DIAGNOSIS OF BREAST CANCER FROM DYNAMIC CONTRAST-ENHANCED MAGNETIC RESONANCE BREAST IMAGES

By

Jacob Levman

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy, Department of Medical Biophysics, University of Toronto.

© Copyright Jacob Levman (2010)
The goal of this research is to improve the breast cancer screening process based on magnetic resonance imaging (MRI). In a typical MRI breast examination, a radiologist is responsible for visually examining the MR images acquired during the examination and identifying suspect tissues for biopsy. It is known that if multiple radiologists independently analyze the same examinations and we biopsy any lesion that any of our radiologists flagged as suspicious then the overall screening process becomes more sensitive but less specific. Unfortunately cost factors prohibit the use of multiple radiologists for the screening of every breast MR examination. It is thought that instead of having a second expert human radiologist to examine each set of images, that the act of second reading of the examination can be performed by a computer-aided detection and diagnosis system. The research presented in this thesis is focused on the development of a computer-aided detection and diagnosis system for breast cancer screening from dynamic contrast-enhanced magnetic resonance imaging examinations. This thesis presents new computational techniques in supervised learning, unsupervised learning and classifier visualization. The techniques have been applied to breast MR lesion data and have been shown to outperform existing methods yielding a computer aided detection and diagnosis system with a sensitivity of 89% and a specificity of 70%.
Acknowledgements

I would like to thank my wife and family, without which the best things I have ever done would not have been possible.

I am thankful for the guidance and advice I have received from my supervisor, Dr. Anne Martel and my advisory committee, Dr. Don Plewes and Dr. Greg Stanisz.

I would also like to particularly highlight the contributions of Dr. Ellen Warner, the head oncologist on the breast MRI screening trials who is responsible for the care of the hundreds of patients enrolled in screening. Furthermore, Dr. Petrina Causer, the head radiologist on the screening trials, is responsible for diagnosing lesions visible on the breast MRI images. Without accurate diagnoses of many small tumours by Dr. Causer, it would be impossible to build an effective computer-aided diagnosis (CAD) system capable of detecting small lesions (which is the main goal of a CAD system). Dr. Don Plewes and his research group have been responsible for ensuring that we have high quality MRI data that facilitates the radiologist detecting cancer in general and small tumours in particular. Additionally, Dr. Anne Martel and her research group was responsible for producing an image registration system that compensates for patient motion that occurs during the examination and has been in use clinically.
# Table of Contents

1. Introduction .................................................. 1
   1.1 Introduction – Breast MRI .................................. 10
   1.2 Introduction – Computer-Aided Detection of Breast Cancer .................. 15
   1.3 Introduction – Classification / Supervised Learning ......................... 19
   1.4 Introduction – High Dimensional Visualization ............................... 22
   1.5 Introduction – Unsupervised Learning / Region-of-Interest Identification .... 23
   1.6 Introduction – Overview of this Thesis .................................. 29

2. Classification: Dynamic Information .......................... 31
   2.1 Introduction – Classification: Dynamic Information ......................... 32
      2.1.1 Methods – Classification: Dynamic Information ......................... 33
      2.1.2 Results – Classification: Dynamic Information ......................... 44
      2.1.3 Discussion – Classification: Dynamic Information ...................... 54

3. Segmentation .................................................. 61
   3.1 Introduction – Segmentation .................................. 62
      3.1.1 Methods – Segmentation .................................. 63
      3.1.2 Results – Segmentation .................................. 77
      3.1.3 Discussion – Segmentation .................................. 82

4. Classification: Dynamic and Shape Information ............... 88
   4.1 Bias-similarity supervised learning .................................. 89
      4.1.1 Methods – Classification: Dynamic and Shape Information .......... 90
      4.1.2 Results – Classification: Dynamic and Shape Information .......... 95
      4.1.3 Discussion – Classification: Dynamic and Shape Information .......... 99

5. Conclusions and Future Work .................................. 106
   5.1 Conclusions .................................................. 107
   5.2 Future Work .................................................. 112

References ................................................... 117
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Breast cancer detection methods and their shortcomings</td>
<td>8</td>
</tr>
<tr>
<td>II</td>
<td>Summary of pattern recognition techniques addressed in this thesis</td>
<td>19</td>
</tr>
<tr>
<td>III</td>
<td>Quantity and Pathological Diagnosis of Breast Lesions</td>
<td>34</td>
</tr>
<tr>
<td>IV</td>
<td>Results of Leave-One-Out Trials – Linear Kernel</td>
<td>45</td>
</tr>
<tr>
<td>V</td>
<td>Results of Leave-One-Out Trials – Polynomial Kernel</td>
<td>45</td>
</tr>
<tr>
<td>VI</td>
<td>Results of Leave-One-Out Trials – Radial Basis Function Kernel</td>
<td>46</td>
</tr>
<tr>
<td>VII</td>
<td>Results of Pixel-by-Pixel Randomized Trials</td>
<td>49</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Example X-ray Mammography Breast Image</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>Example Ultrasound Breast Image</td>
<td>6</td>
</tr>
<tr>
<td>1.3</td>
<td>Example PET Image of Metastasized Breast Cancer</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>A Two-Compartment Pharmacokinetic Model</td>
<td>9</td>
</tr>
<tr>
<td>1.5</td>
<td>Example Breast MR Subtraction Images</td>
<td>12</td>
</tr>
<tr>
<td>1.6</td>
<td>Example Signal Intensity Time Curves</td>
<td>15</td>
</tr>
<tr>
<td>1.7</td>
<td>Block Diagram for Final Computer-Aided Detection System</td>
<td>18</td>
</tr>
<tr>
<td>1.8</td>
<td>Example Two-Class Data with Separating Decision Function</td>
<td>20</td>
</tr>
<tr>
<td>1.9</td>
<td>The Role of the Enhancement Threshold (Enh. Thr.) in CAD</td>
<td>26</td>
</tr>
<tr>
<td>1.10</td>
<td>Example Signal Intensity Curves with the Enhancement Threshold</td>
<td>27</td>
</tr>
<tr>
<td>2.1</td>
<td>Classifier Visualization Example</td>
<td>42</td>
</tr>
<tr>
<td>2.2</td>
<td>Classifier Visualization on Leave-One-Out Data</td>
<td>47</td>
</tr>
<tr>
<td>2.3</td>
<td>Classifier Visualization on Pixel-By-Pixel Data</td>
<td>50</td>
</tr>
<tr>
<td>2.4</td>
<td>Sampled Signal Intensity Time Curves from figure 2.3</td>
<td>51</td>
</tr>
<tr>
<td>2.5</td>
<td>Invasive and Non-invasive Cancers in Principal Component (PC) Space</td>
<td>51</td>
</tr>
<tr>
<td>2.6</td>
<td>An Example Malignancy Diagnosed as Cancerous by SVMs</td>
<td>52</td>
</tr>
<tr>
<td>2.7</td>
<td>A Malignant Lesion Diagnosed at Different Enhancement Thresholds</td>
<td>53</td>
</tr>
<tr>
<td>3.1</td>
<td>A Hierarchical Approach to Segmentation</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Pane Merger Process</td>
<td>65</td>
</tr>
<tr>
<td>3.3</td>
<td>Multidimensional to Unidimensional Projection</td>
<td>69</td>
</tr>
<tr>
<td>3.4</td>
<td>Block Diagram: Atypical Segmentation Evaluation Methodology</td>
<td>70</td>
</tr>
<tr>
<td>3.5</td>
<td>Effect of the Input Parameter on Resulting Segmentations</td>
<td>78</td>
</tr>
<tr>
<td>3.6</td>
<td>ROC Areas for each Feature Measurement</td>
<td>79</td>
</tr>
<tr>
<td>3.7</td>
<td>Best Performing ROC Areas at a Fixed Segmentation Parameter Setting</td>
<td>80</td>
</tr>
<tr>
<td>3.8</td>
<td>An MR Image of a Segmented Malignant Lesion with Diffuse Edges</td>
<td>81</td>
</tr>
<tr>
<td>4.1</td>
<td>Behaviour of Proposed Classifier</td>
<td>92</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparative Supervised Learning Validation Results</td>
<td>97</td>
</tr>
<tr>
<td>4.3</td>
<td>PC Plots Showing SVMs and Proposed Supervised Learning Technique</td>
<td>98</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>4.4 PC Plots Showing SVM Shortcomings</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>5.1 A Montage of Example Segmentations</td>
<td>109</td>
<td></td>
</tr>
</tbody>
</table>
List of Symbols and Abbreviations

2D Two-Dimensional
3D Three-Dimensional
4D Four-Dimensional
ANN Artificial Neural Network
BI-RADS Breast Imaging-Reporting and Data System
CAD Computer Aided Diagnosis and Detection
CADe Computer Aided Detection
CADx Computer Aided Diagnosis
CT Computed Tomography
DCE-MRI Dynamic Contrast-Enhanced Magnetic Resonance Imaging
DCIS Ductal Carcinoma in Situ
ErR The Error Penalty Ratio
FDG Fluorodeoxyglucose
FOS Fast Orthogonal Search
FOV Field of View
FPP False Positive Pixels
Gd-DTPA Gadolinium diethylenetriamine penta-acetic acid – contrast agent
IDC Invasive Ductal Carcinoma
MRI Magnetic Resonance Imaging
NPV Negative Predictive Value
OA Overall Accuracy
PC Principal Component
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Principal Components Analysis</td>
</tr>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
<tr>
<td>RBF</td>
<td>Radial Basis Function</td>
</tr>
<tr>
<td>RGB</td>
<td>Red Green Blue</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristic</td>
</tr>
<tr>
<td>ROI</td>
<td>Region-of-interest</td>
</tr>
<tr>
<td>SER</td>
<td>Signal Enhancement Ratio</td>
</tr>
<tr>
<td>SI</td>
<td>Signal Intensity</td>
</tr>
<tr>
<td>SPGR</td>
<td>Spoiled Gradient</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>$T_1$</td>
<td>Longitudinal relaxation time</td>
</tr>
<tr>
<td>$T_2$</td>
<td>Transverse relaxation time</td>
</tr>
<tr>
<td>TE</td>
<td>Excitation Time</td>
</tr>
<tr>
<td>TPP</td>
<td>True Positive Pixels</td>
</tr>
<tr>
<td>TR</td>
<td>Recovery Time</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
In Canada, breast cancer has the third highest incidence rate and the third highest mortality rate of any type of cancer [1]. In women, who account for 99% of breast cancer cases in Canada, the disease has the highest incidence rate and second highest mortality rate (after lung) of any cancer type [1]. Because of excellent treatment options for small tumours, early detection has been identified as key to improving survival rates for this disease [2].

Whether early detection by existing breast cancer screening does in fact reduce the mortality rate of the disease is a subject of some controversy. L. Bonneux demonstrates that mammographic screening reduces mortality [3]. Alternatively, H. Hewitt argues that screening for breast cancer has no impact on mortality [4], however, his analysis was made in 1993 and looked exclusively at the performance of the existing technologies of the day (x-ray mammography, the only modality that has been tested for mortality). Furthermore, treatment options for cancerous tumours have been improving, which reinforces the benefit obtained from early detection. One of the main arguments in favour of breast cancer screening is that since the incidence rate of breast cancer has been increasing in the industrialized world, a combination of screening and effective treatments must be responsible for preventing the mortality rate from increasing. These arguments have focused on traditional breast screening (by x-ray mammography) which has been the standard imaging modality for over 30 years. Incidence rates have been increasing, but this is expected given that many people are screened by methods that didn’t exist a few decades ago. It should be noted however, that in Canada, net mortality due to breast cancer began falling around 1990 [1], a period where Canada was growing
in terms of population. It has also been demonstrated that x-ray mammography has lead to a 15-45% reduction in the mortality rate [5, 6].

Both x-ray imaging technologies and tumour treatment options have continued to improve in recent years. A more recent thorough retrospective analysis performed in Denmark (published in 2005) indicates that breast cancer screening by mammography results in a 25-37% reduction in the mortality rate (depending on how many people volunteer for screening) [7]. Alternative breast cancer screening methods have become more common in the last 15 years (Ultrasound, Magnetic Resonance Imaging - MRI, Positron Emission Tomography - PET), however, it is yet to be determined if newer techniques with greater sensitivity (like MRI) will result in breast cancer mortality improvements. In the research based screening program run at Sunnybrook hospital many small (2-3 mm) lesions have been identified with MRI. Since the cure rate for small cancers that have not yet invaded neighbouring tissues is now about 99%, it is plausible that any imaging technology that allows the detection of small cancers will assist in reducing the mortality rate of the disease.

Several techniques exist to assist in the process of screening for breast cancer. Palpation is one of the most common methods for breast cancer screening and is typically performed either as a self examination or clinically by a trained physician. Palpation involves the manual inspection of the breast to find any hard malignant-like lumps. While the technique can be performed at home or by any medical doctor, the technique’s
X-ray mammography is the most common imaging method for the detection of breast cancer due to its affordability and the high availability/accessibility of the technology for the past 30 years. An example mammographic image is provided in Fig. 1.1. In mammography, x-ray projection images of the breasts are acquired. Tumours are imaged because they tend to be comprised of dense malignant tissues which absorb x-rays. However, x-rays are also absorbed by healthy fibroglandular tissue which can significantly limit the ability of the test to detect tumours. As such, it is known that the technique has low sensitivity (62-69%) for women with dense breasts. Furthermore, according to one large study (463,372 examinations) dense breasts are quite common, representing 44% of the total set of examinations [8]. X-ray mammography improvements are being researched based on tomography [9] (where a series of projection images at different angles are acquired for each breast). Another area of research in x-ray mammography is the use of contrast agents (e.g. iodine) injected into the patient’s blood stream that absorbs x-rays and thus assists in making tumours visible on the resultant images [10].
Another common method for breast cancer detection is ultrasound. Ultrasound involves detection of the reflection of a sound wave off of a tumour. Ultrasound can have difficulty differentiating a malignancy from some types of non-malignant tumours (like fibroadenomas). However, ultrasound is particularly useful in determining cysts (which are often fluid filled and appear dark and uniform on ultrasound images) to be non-malignant. An example ultrasound image with a cyst is provided in Fig. 1.2. Research is being conducted on improving ultrasound based breast cancer screening with the use of
microbubble contrast agents which are small gas filled bubbles with a high degree of echogenicity (the ability of an object to reflect ultrasound waves) [11].

![Image of a breast ultrasound image of a cyst](http://rad.usuhs.edu/medpix/).

**Fig. 1.2.** An example breast ultrasound image of a cyst (obtained from online database [http://rad.usuhs.edu/medpix/]).

Both positron emission tomography (PET) and magnetic resonance imaging (MRI) based cancer screening involve the injection of a contrast agent which pools into the lesion tissue. Cancer is characterized by cells that are growing in an uncontrolled manner. These cells release signaling molecules (like vascular endothelial growth factor - VEGF) that indicate that they need to grow and need more nutrients. These signaling molecules trigger nearby blood vessels to undergo angiogenesis – the formation of new blood vessels. These new blood vessels tend to be characteristically leaky and provide a pathway for the delivery of our injected contrast agent to the site of the malignancy. In PET, a radiopharmaceutical tracer (such as FDG – fluorodeoxyglucose) is injected into
the blood stream and leaks into tumour tissues. FDG is a glucose analog and therefore accumulates preferably in the areas of high glucose metabolism (i.e. cancer cells). The radiopharmaceutical undergoes decay, thus emitting radiation. This radiation is detected to form an image of the patient. These nuclear medicine techniques can be quite useful in detecting late stage cancer that has spread to other tissues and organs as is illustrated in Fig. 1.3. As in PET, MRI involves the injection of a contrast agent into the blood stream and also pools in tumour tissues. An MRI-based contrast agent acts as a little magnet with a dipolar moment that assists nearby hydrogen protons to realign with the main magnetic field (thus reducing their relaxation time).

Fig. 1.3. An example PET image of breast metastases (obtained online http://www.petscaninfo.com/)

As previously introduced, several techniques exist to assist in the process of screening for breast cancer. Table I is provided to summarize the main methods available for screening.
Table I: Summary of main breast cancer detection methods and their shortcomings

<table>
<thead>
<tr>
<th>Screening Method</th>
<th>Description</th>
<th>Shortcomings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast examination</td>
<td>Breast tissue is palpated to find hard tumour tissue</td>
<td>Low sensitivity</td>
</tr>
<tr>
<td>(clinical or self)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mammography</td>
<td>X-rays are absorbed in a typically dense malignancy</td>
<td>X-rays are also absorbed by natural fibroglandular tissue Poor detail on dense breasts</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>Sound waves are reflected off tumour</td>
<td>Better for detection of non-malignant cysts than malignancies</td>
</tr>
<tr>
<td>MRI</td>
<td>Contrast agent pools in leaky tumour enhancing signal from local protons</td>
<td>High variability in radiological analysis of large datasets</td>
</tr>
<tr>
<td>PET / nuclear imaging</td>
<td>Contrast agent pools in leaky tumour emitting radiation that is detected</td>
<td>Better for detection of late stage metastasized breast cancer than localizing small tumours within the breast</td>
</tr>
</tbody>
</table>

When a contrast agent is injected into the blood stream, it will circulate and then reach the tumour site. Because the blood vessels feeding the tumour are characteristically leaky, the contrast agent will pool into the lesion until its concentration within the lesion exceeds its concentration in the blood stream, at which point it begins to diffuse back into the vasculature. Considerable research has been conducted on modeling this contrast agent behaviour in a process known as pharmacokinetic modeling. A typical pharmacokinetic approach will model the concentration of contrast agent in the blood plasma (Cp), the permeability of the vasculature (Ktrans) and the concentration of the contrast agent in the lesion’s extracellular space (Ce). This is the commonly used two-compartment model (see Figure 1.4) used in the most common pharmacokinetic approaches [12, 13]. Some pharmacokinetic approaches model even more details like the concentration of contrast agent within the cells and could thus be a three compartment model or more [14].
In the context of a computer-aided diagnostic system for breast MRI, a pharmacokinetic model can allow the standardization of any signal intensity time curve into measurements which include the permeability of the vasculature ($K_{\text{trans}}$). $K_{\text{trans}}$ has diagnostic potential as we know cancerous tumours tend to be fed by characteristically leaky angiogenic blood vessels. In practice, taking a signal intensity time curve and fitting it to a pharmacokinetic model can result in a loss of some of the relevant information that helps to discriminate between cancerous and benign lesions. This is because the curve fitting process is subject to error. Additionally, the curve fitting process becomes less reliable as the number of MRI samples is reduced. The screening data used in this thesis was generated with 4 post-contrast MRI acquisitions, which is an inadequate amount for performing reliable curve fitting, thus pharmacokinetic modeling was deemed inappropriate for this research project.
1.1 Introduction – Breast MRI

Large scale breast cancer screening trials have been ongoing at Sunnybrook Health Sciences Centre. This screening program is the largest, single institution study of Breast MRI ever performed. The screening program compares the performance of dynamic contrast enhanced magnetic resonance imaging (DCE-MRI), ultrasound, mammography and clinical breast examination as diagnostic screening methodologies for high risk women [15]. This thesis would not have been possible if it were not for the efforts of dozens of research staff (oncologists, radiologists, pathologists, MRI technicians, research coordinators, imaging physicists and more) to support the ongoing breast cancer screening program. Between November 3, 1997 and August 21, 2008, 550 high risk women were recruited from familial cancer clinics in southern Ontario and Montreal, Canada. Participation in screening was offered to all eligible women in the context of genetic counseling. Informed consent was obtained from all participants. The retrospective analyses presented in this thesis were approved by the institutional review board of Sunnybrook Research Institute. Annual screening of the patient population has resulted in 1749 breast MR examinations.

Although this has been the largest single-institution breast screening trial, prospective screening trials have been in place in many different centres around the world [16-19]. Results from these studies indicate that magnetic resonance imaging is the most sensitive modality compared.
Genetic mutations on the BRCA1/2 genes are estimated to produce an 85% lifetime risk of developing breast cancer [20]. It has been demonstrated that in three percent of women diagnosed with unilateral breast cancer by mammography, a pre-treatment MRI examination detects mammographically occult cancer in the contralateral breast [21-23]. These studies suggest that MRI has a promising future in improving the clinical diagnosis of breast cancer. A significant barrier for the widespread introduction of MRI based breast cancer screening is the cost of screening (approximately 10 times as much as x-ray mammography [24]). However, new technologies that improve the throughput of patients being screened and facilitate the biopsy of suspicious tissues have been developed that reduce the overall cost of MRI based screening to 4 times that of mammography (Sentinelle Medical Inc., Toronto, ON, Canada, US Patent Application No. 12277061) [25]. Example breast MRI subtraction images are provided in Fig. 1.5.
Fig. 1.5. Subtraction images of a malignant invasive ductal carcinoma (top) and a benign fibroadenoma (bottom).

The main method of screening for breast cancer with MRI involves acquiring a volumetric T$_1$ weighted image (a set of 2D images/slices lined up next to each other). The patient is then injected with a contrast agent and consecutive volumetric images are acquired over time. Since cancerous tumours are typically fed by characteristically
leaking angiogenic blood vessels, the injected contrast agent accumulates in potentially malignant lesions. A radiologist is typically responsible for visually inspecting the images and predicting malignancies based on how the lesion’s brightness changes with time, and based on the shape of the lesion. This forms the most common method for detecting breast cancer from MRI, however, alternative magnetic resonance based techniques are also available. The above technique involves T₁ imaging - whereby the MRI signal that forms the image is acquired from hydrogen protons precessing in the z plane. Alternative techniques for breast cancer detection from MRI involves T₂ imaging – whereby the MRI signal that forms the image is acquired from hydrogen protons precessing in the x-y plane [26].

Breast cancer can also be detected with magnetic resonance imaging in unusual ways. Quantitative perfusion measurements, such as the apparent diffusion coefficient (where the diffusion of hydrogen is measured) can also be used to detect cancer by MRI [27]. Cancerous tumours tend to have a small extracellular space due to the proliferation of malignant cells causing a reduction in the amount of measured diffusion in the overall lesion. Diffusion measurements are sensitive to tissue microstructure but are not very specific so are not widely used. Magnetic resonance spectroscopy is another method for the detection of breast cancer, whereby the imaging technique measures an elevated choline peak [28]. Choline plays a role in cell metabolism and so it is expected that metabolically active malignant cells will exhibit higher signal to noise ratios for choline. Sodium is also known to play a role in cell metabolism and MRI has been investigated as a method for imaging sodium which may assist in the monitoring of a malignant lesion’s
therapeutic response [29]. The signal to noise ratio of sodium imaging is low and so the technique is not widely used. Additionally, magnetic resonance elastography can be used as it measures tissue hardness, an effect similar to palpation but unlike palpation the measurement can be made from anywhere in the breast. The technique is performed by stimulating the breast with an ultrasound transducer and synching the MRI pulse sequence to the ultrasound signal [30]. Although many MRI based techniques exist, traditional T\textsubscript{1} weighted imaging before and after the injection of a contrast agent is by far the most established method. It should be noted, however, that even in this method there are considerable variations in terms of the pulse sequence used for image acquisition; this can significantly affect the temporal and spatial resolution as well as the signal-to-noise ratio (SNR) of the breast MRI examination images and therefore influence MRI specificity and sensitivity in cancer detection. This thesis is focused on improving the most common type of MRI based breast cancer screening: T\textsubscript{1} weighted dynamic contrast-enhanced magnetic resonance imaging followed by a Gd-DTPA injection and more T\textsubscript{1} weighted imaging.

A breast MRI examination involves the injection of a contrast agent which causes tissues to enhance or brighten. An expert radiologist or a computer-aided detection and diagnosis system will then identify suspected malignant lesions by examining how the brightness (signal intensity enhancement) changes over the course of the examination. Expected brightness time patterns are provided in Fig. 1.6. An expert radiologist or a computer-aided diagnosis and detection (CAD) system can also identify suspected malignant lesions by examining the physical shape of the enhancing lesion. A radiologist’s breast
MRI lesion classification is standardized with BI-RADS (Breast Imaging-Reporting and Data System) and a lesion’s description is standardized in the breast MRI lexicon [31]. Standard reporting descriptions include whether the lesion is a focal enhancement, the appearance of the lesion’s margin, the shape of the overall lesion, the enhancement pattern of the lesion, the enhancement pattern of non-mass tissues (ie. ducts connected to the lesion etc.) and observations of breast symmetry in bilateral studies.

![Graph](image)

**Fig. 1.6.** An example cancerous (solid) and non-cancerous (dashed) signal source.

### 1.2 Introduction – Computer-Aided Diagnosis/Detection of Breast Cancer

Although MRI has been shown to be an effective modality for breast cancer screening [15], there is considerable inter-observer variability between radiologists in their interpretation of the large amounts of data acquired in a breast MRI examination. In the United Kingdom a major study has been devoted to this question [32]. Multi-centre screening trials have been ongoing in the United Kingdom, and this study looked at the diagnostic variability between 15 different radiologists all analyzing the same breast MRI examinations. The sensitivity of cancer detection ranged between the radiologists from
77.4% to 94.7%. The specificity of the diagnoses ranged between radiologists from 60.5% to 76.7%. It is known that if multiple radiologists independently analyze the same examinations, and if we were to biopsy any lesion that these radiologists flagged as suspicious, then the overall screening process would become more sensitive [33] (but less specific). Unfortunately, it is prohibitively expensive for multiple radiologists to be involved in the screening of every breast MRI examination. It is hypothesized that instead of having a second expert human radiologist examine each set of images, a second reading of the examination could be performed by a computer-aided detection and diagnosis system. Computer-aided diagnosis and detection (CAD) systems have the potential to further improve MRI based breast cancer screening by reducing inter-observer variability and potentially standardizing all radiologists to the best sensitivity and specificity values possible. The research presented in this thesis is focused on the development of a computer-aided detection and diagnosis system for breast cancer screening from dynamic contrast-enhanced magnetic resonance imaging examinations.

The field of computer-aided diagnosis and detection (CAD) of breast cancer using MRI builds on previous research in CAD applied to breast cancer detection from mammography and ultrasound. Pattern recognition CAD techniques similar to those used in this thesis have been previously researched for their potential use in CAD x-ray mammography systems [34, 35] and CAD ultrasound systems [36]. X-ray mammography and ultrasound CAD systems are developed to analyze the types of images generated by those respective modalities. The nature of the image data acquired through x-ray mammography or ultrasound is very different from the image data acquired in a breast
MRI examination and a radiologist will look for features (like lesion enhancement) in a breast MRI examination that would not be available in a typical mammography or ultrasound based exam. Overall CAD system performance is always limited by the ability of the imaging modality to clearly delineate between a cancerous tumour and surrounding tissues. CAD performance is also limited by the imaging modality’s ability to clearly delineate between cancerous tumours and non-cancerous tumours found at different image locations or on different examinations. Improvements to image contrast or resolution will likely assist the overall screening process (and CAD detection and diagnosis process) for any of the given modalities.

In order to further improve the DCE-MRI based breast cancer screening process significant research has been conducted towards developing a computer aided detection and diagnosis system to assist radiologists in image analysis. Given that the aforementioned screening trials at Sunnybrook Health Sciences Centre [15] employ an imaging protocol that generates five three dimensional (3D) volumes each containing 28-32 slices per breast, a motivating factor for the development of a computer-aided diagnostic and detection system is to reduce the complexity of the data for radiological analysis. A key component of such a computer-aided diagnosis and detection system is an appropriate segmentation algorithm responsible for identifying lesions that could represent malignancy with a region-of-interest (computer-aided detection – CADe). A second key component of such a CAD system is an appropriate classification algorithm responsible for making a final decision as to whether suspect tissue represents a malignant or benign lesion (computer-aided diagnosis – CADx).
There are a multitude of computational techniques for automatically detecting breast cancer from MRI. In an effort to improve the breast MRI screening process a number of new statistical and machine learning techniques have been developed to assist in computer-aided detection and diagnosis. Each of the new types of computational techniques addressed in this thesis will be introduced individually: supervised learning (classification), classification visualization (evaluation) and unsupervised learning (segmentation / clustering / region-of-interest (ROI) identification). The techniques come together to form an overall computer-aided detection and diagnosis system as represented by Fig. 1.7.

![Fig. 1.7. Block diagram for final computer-aided detection and diagnosis system](image)
This thesis covers three main pattern recognition topics: unsupervised learning, supervised learning and the evaluation of a supervised learning technique. Table II provides a summary and introduction of each of these techniques, and each will be introduced individually and in more detail in the following sections.

Table II: Summary of pattern recognition techniques addressed in this thesis

<table>
<thead>
<tr>
<th>Description</th>
<th>Unsupervised Learning</th>
<th>Supervised Learning</th>
<th>Evaluation of a supervised learning technique</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Breast MRI Context</strong></td>
<td>Identify regions-of-interest on an image: no pre-existing knowledge of data</td>
<td>Identify specific regions with pre-existing knowledge of data</td>
<td>Allows us to compare relative performance of different supervised learning techniques</td>
</tr>
<tr>
<td><strong>Existing Methods and their Shortcomings</strong></td>
<td>Divides exam into potentially malignant regions</td>
<td>Clarifies potential malignancy as cancer or not</td>
<td>Selecting an appropriate supervised technique for breast cancer detection</td>
</tr>
<tr>
<td>c-means / k-means</td>
<td>Support Vector Machines</td>
<td>Leave-one-out validation</td>
<td></td>
</tr>
<tr>
<td>● Pre-selecting number of regions is inappropriate for screening</td>
<td>● Two parameters to tune to get best performance</td>
<td>● Can give unrealistic performance measures</td>
<td></td>
</tr>
<tr>
<td>Markov random field</td>
<td>Linear Discriminant Analysis</td>
<td>Randomized validation</td>
<td></td>
</tr>
<tr>
<td>● Many parameters to tune to get appropriate regions</td>
<td>● Underperforms SVMs</td>
<td>● Lots of computation time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>● Can be difficult to compare lots of numbers</td>
<td></td>
</tr>
</tbody>
</table>

1.3 Introduction – Classification / Supervised Learning

These are a class of pattern recognition techniques that are formal mathematical rules that govern how to predict which group a sample belongs to based on a set of labeled training data. This prediction process is performed by a computer which implements the formal mathematical rule. This allows a researcher to collect a set of measurements from many samples with a known gold standard and use the information to classify future unknown
cases. In this thesis these types of techniques are being applied to distinguish malignancies from non-malignant tissues. This is accomplished by collecting a training set of MRI data representing malignant and benign breast lesions. For each lesion a set of measurements are made (ideally that help distinguish malignancies, such as lesion irregularity or lesion enhancement) and these measured values along with the class information (malignant or benign) is provided to a machine learning technique as the training data. This algorithm will then be responsible for predicting malignancies on new breast MRI lesions whose diagnosis is not known to the computer program. An example is provided in Fig. 1.8, whereby the two training groups (represented by the red and green dots) have been provided to a machine learning algorithm which has then selected the solid black line as a decision mechanism for making future predictions on samples with no known class.

**Fig. 1.8.** Example two-class (and two measurement) training data (red and green dots) separated by a machine learning decision function (solid black line) to be used to make future decisions.
Substantial research has been conducted on the application of computer classification methods (supervised learning) to the analysis of breast MRI. Artificial neural networks (ANNs) have been one of the most common approaches for researching the classification of malignant and benign breast MRI lesions [37-43]. An ANN is a computer algorithm that mimics the behaviour of a real neural network (one found in the brain). Unfortunately, typical ANN implementations involve far fewer artificial neurons than would be involved in the same prediction process as implemented in a brain. Artificial neural networks have not been used in this thesis as there is infinite flexibility in the implementation of the ANN and there is no accepted gold-standard ANN architecture for breast MRI. Statistical based alternatives to ANNs have been investigated in this thesis. Linear discriminant analysis, a classical classification approach, has also been applied to breast MRI analysis [44, 45], however it has been shown to underperform support vector machines [46, 47] which can model nonlinear class boundaries. Significant research has been conducted on a breast MRI lesion’s signal intensity time curve’s wash-in and wash-out (kinetic) characteristics [44, 48-50]. Some approaches have led to commercial classification aids in use by radiologists (Confirma Inc., Kirkland, WA, USA - Sentinelle Medical Inc., Toronto, ON, Canada), however, these techniques have been shown to under-perform the statistical based support vector machine approach [51].

Support vector machines (SVMs) have been shown to perform well as a computer-aided diagnostic classification mechanism for breast cancer screening in ultrasound [36] and mammography [34]. More recently it has been shown that SVMs outperform a variety of other machine learning techniques when applied to the separation of malignant and
benign DCE-MRI breast lesions [47, 52]. Support vector machines operate by locating a line (or surface) that attempts to split the training data into two categories. This surface is selected such that its distance (margin) to the nearest training data on either side of the surface is maximized. More information on support vector machines is provided in sections 2.1.1 and 4.1.1.

1.4 Introduction – Classification Evaluation / Visualization

In addition to classification research, this thesis is also concerned with the evaluation of any given supervised learning technique (classifier – introduced in section 1.2) through visualization techniques. In a typical situation, a classification technique is evaluated by performing a plethora of validation based calculations. The large sets of numbers computed are then compared to determine which technique is going to perform the best and is the most reliable. Typically many tables of calculations need to be compared in order to draw firm conclusions about the quality of any given classification technique. This thesis involves the presentation of a new visual based evaluation method for a classification technique.

It is normal in classification research to have many measurements per sample. In this situation, the visual evaluation of a classification function can be very challenging. Techniques exist (like principal components analysis) to rotate high dimensional data such that the researcher can visually observe a projection of the data on a standard two-dimensional plot. This thesis demonstrates a novel method for projecting a classification function onto a two-dimensional projection of high-dimensional data. This allows for
easy visual interpretation of the behaviour of a classification function with respect to our data. More detail on the proposed technique is provided in section 2.1.1.

Some related research has been conducted on this topic. Nattkemper and Wismuller have researched the use of self organizing maps for tumour feature visualization from breast MRI data [53]. Komura et al. have proposed multidimensional support vector machines as a mechanism for the visualization of high dimensional data sets and demonstrated the technique with gene expression data [54]. Somorjai and Dolenko have proposed the visualization of high dimensional data onto a special plane called the relative distance map [55]. All of these techniques are methods for projecting high-dimensional data onto an easy-to-view two-dimensional plot. However, the visualization technique presented in this thesis can be thought of as an extension that allows the researcher to project a classification function (the boundary that makes decisions like whether a given tissue is representative of cancer) onto an easy-to-view two-dimensional data plot.

1.5 Introduction – Unsupervised Learning / Region-of-Interest Identification

The research presented in this thesis also covers the topic of unsupervised image segmentation – also called region-of-interest identification, clustering or unsupervised learning. Image interpretation forms a critical step in a variety of fields including medical imaging, remote sensing / satellite imaging, robotics, computer vision, security / facial recognition etc. Identifying salient regions-of-interest (ROIs) in images often forms a crucial first step in the overall interpretation of that image. In medical imaging, the identification of ROIs is a common step towards identifying abnormal tissues and
monitoring a patient’s treatment, a good example being detecting breast cancer lesions from magnetic resonance images (MRI) [56] which are known to be more likely to be malignant if the ROI is non-spherical. In remote sensing the identification of ROIs from satellite images can facilitate the monitoring of deforestation [57] and other forms of land use and for monitoring weather patterns. In security and computer vision applications, accurate ROIs can facilitate the identification of a particular person through techniques such as facial recognition or by analyzing an individual’s retinal image [58]. ROIs are also useful in robotics, where an autonomous robot identifies particular objects in its field of view in order to perform a particular task [59]. This thesis presents a new technique for identifying regions-of-interest in any type of image (see chapter 3).

Considerable research emphasis has been placed on using advanced techniques to assist in the identification of an image’s salient regions-of-interest. The relevant literature often refers to this process as either image segmentation or clustering. The bulk of current research in this field involves using existing techniques (sometimes modified for the target application) to identify ROIs for a very particular application or image type [56-60]. These approaches, while producing useful results for the application at hand, are often overspecific to the target application and as such will not necessarily be a useful technique once the image type or target application has changed. One of the largest barriers for the widespread use of existing techniques is that many require the user/researcher to correctly set many input parameters in order to obtain the desired ROIs.
Although many image segmentation (ROI identification) techniques exist, it is common for these segmentation algorithms to behave unpredictably when presented with an image type not previously tested. However, it is known that ROI identification is performed efficiently and effectively in the human mind. As such, methods for mimicking the image interpretation that takes place naturally in the brains of humans were pursued for this thesis. In 2004 Jeff Hawkins published a theory of intelligence [61]. This theory posits that the problem solving abilities of the human brain are largely accomplished in the neocortex. The theory supposes that the brain and particularly the neocortex is comprised of collections of neurons each of which implements the same basic learning method repeated over and over again in a hierarchical fashion. An illustrative description of a repetitive hierarchical algorithm is available in section 3.1.1. In 2007 the first computer algorithm was developed based on this theory of intelligence and made available to the public (www.numenta.com). The original system was developed towards the goal of creating computer programs that identify images that are similar to ones the machine had previously been trained on (called supervised learning in the literature). Effectively, the machine is initially instructed what it is looking at and can predict the content of a new image based on this experience. However, this is different from the task of automatically identifying regions-of-interest in an image when the computer has not been provided with any prior information regarding the types of images it needs to interpret (called unsupervised learning in the literature). This thesis presents a method for automatically identifying regions-of-interest in an image (see chapter 3) without prior knowledge of the type of image data it needs to interpret. This approach is based on the hierarchical agglomerative approach suggested in “On Intelligence” [61], however, in this thesis this
hierarchical approach is being applied to the segmentation process as opposed to the classification process pursued by Numenta (www.numenta.com). It should be noted that this theory leads us to a motivation for selecting a particular class of existing methods (known in the computer science literature as hierarchical agglomerative techniques) as a framework for building our ROI identification system. This thesis applies new statistical metrics that minimize the assumptions we need to make about the data we are segmenting. These metrics are performed at each step in the hierarchy and the overall segmentation process (see chapter 3) represents a novel method for ROI identification.

The most common method for segmenting a lesion in a breast MRI examination is by use of the enhancement threshold which has been used in many studies [41-42,48,62-73]. An example figure depicting the role the enhancement threshold plays in a typical breast MRI CAD system is provided in Fig. 1.9, whereby a tissue location that does not exceed the enhancement threshold is assigned a benign diagnosis.

![Block diagram illustrating the role of the enhancement threshold on the time brightness curve (signal intensity – time curve) in breast MRI CAD systems.](image)

**Fig. 1.9.** Block diagram illustrating the role of the enhancement threshold on the time brightness curve (signal intensity – time curve) in breast MRI CAD systems.
The effect of the enhancement threshold is further illustrated in Fig. 1.10, whereby four example signal-intensity time-curves are provided, two of which exceed the enhancement threshold and two of which do not. The enhancement threshold is used for region-of-interest (ROI) identification (or segmentation) by assigning each spatially contiguous region that exceeds the enhancement threshold as an independent ROI. The enhancement threshold is set in the computer-aided detection and diagnosis system and typically can be controlled by the radiologist using the software. If the threshold is set too high, malignant tissues will be forcibly labeled as benign, if the threshold is set too low, many non-malignant tissues will be CADx tested as a potential malignancy.

**Fig. 1.10.** Example signal intensity time curves with respect to the enhancement threshold. Curves C and D are below the threshold and therefore will both be labeled benign by the CADx system.
Another common approach to the segmentation of breast MRI lesions is to use artificial neural networks [37, 42, 73] and the related self organizing map [74], to assist in the separation of malignant and benign breast MRI lesions. Artificial neural networks attempt to solve this computational problem by mimicking the assumed behaviour of neurons in the brain. Artificial neural networks have not been used in this thesis due to their infinite flexibility in implementation and the lack of a recognized gold standard architecture for breast MRI. Additionally, Stoutjesdijk et al. have applied the mean-shift algorithm to the analysis of breast MRI lesions [75], a technique that groups similar tissues together based on both similar signal intensities (pixel values) and similar locations within the image. Mean-shift was deemed inappropriate for our studies as our lesion population consists of both large and very small tumours which is challenging for the mean-shift algorithm as it has a difficult time segmenting lesions of widely varying size. Breast MRI lesion segmentation has also been performed by 4D co-occurrence texture analysis [57], a technique that allows the identification of patterns of changing signal intensities with varying spatial positions on the images. Chen and Giger have proposed a breast MRI segmentation approach that attempts to compensate for the MRI bias field [76], a common imaging artifact. Chen has also proposed a breast MRI segmentation approach using fuzzy c-means clustering [77]. However, semi-automatic techniques requiring the radiologist to identify a suspect lesion prior to computer-aided diagnosis are still susceptible to human error: the radiologist could overlook a small lesion and so the CADx system would not have the chance to correctly label the potential malignancy.
1.6 Introduction – Overview of this Thesis

The goal of this thesis is to develop computational and pattern recognition techniques for use in a computer-aided detection and diagnosis system for identifying malignant regions on breast MRI examinations. This thesis subsection is intended as a roadmap to help guide the reader through the document. In this thesis, chapter 2 addresses the computer-aided diagnosis of breast cancer from MRI examinations, with segmentation limited to looking at single-pixel regions (thus we are looking at the dynamic information: how the lesion’s brightness changes with time). The chapter covers the use of support vector machines, presents a new method for visually evaluating classification functions and examines the use of the enhancement threshold in breast MRI CADx. Chapter 3 addresses a new technique for image segmentation (identifying regions-of-interest from an input image without prior information). The unsupervised learning technique is applied to volumetric segmentation of breast MRI examinations towards the identification of potentially malignant lesions by facilitating the extraction of regional measurements (such as shape irregularity). It was demonstrated to outperform the widely used enhancement threshold based segmentation. Chapter 4 presents a new bias-similarity based classification (supervised learning) algorithm and demonstrates the advantages the technique offers over the established support vector machine method, as well as how the technique facilitates the commonly used receiver operating characteristic curve based evaluative analysis [78]. The final contribution of this thesis is to demonstrate how to apply the aforementioned statistical and computational techniques towards the detection of breast cancer from dynamic contrast-enhanced magnetic resonance imaging examinations combining both shape and dynamic information. In this
system an entire input breast MRI examination is segmented (or parcellated) by the proposed segmentation technique (chapter 3), a set of relevant measurements are made and final predictions of malignancy are based on a pre-collected set of these measurements (training data) from malignant and benign lesions using the proposed classification technique (chapter 4). Chapter 5 addresses the final conclusions of this thesis and recommendations for future work.
Chapter 2

Classification: Dynamic Information

Some of the material presented in this chapter has been published in the journal *IEEE Transactions on Medical Imaging* [51]. Some of the material presented in this chapter has been published with the journal *Academic Radiology* [79]. Some of the research presented in this thesis chapter won 3rd prize in the 5th Annual Imaging Symposium Poster Competition at the Imaging Network Ontario Symposium in Toronto, 2006. Some of this research was also presented at the MICCAI 2006 Workshop on Medical Image Processing, Copenhagen, Denmark [47].
This thesis chapter addresses the classification of breast cancer from MRI examinations based on the dynamic information: how a suspect lesion’s brightness changes over the course of the examination. A typical CAD system will involve a segmentation phase prior to this classification phase that detects potential malignancies. In this chapter, the segmentation phase is simplified such that each pixel/voxel in an examination is considered an independent region-of-interest. Thus this chapter is concerned with diagnosing cancer based only on the dynamic information (how a tissue’s brightness changes over the course of the examination). The segmentation process, whereby neighbouring similar tissues are grouped together for combined analysis is addressed in chapter 3. The material presented in this chapter is intended to thoroughly analyze the use of support vector machines for the computer-aided diagnosis of breast cancer from MRI. This chapter also presents a novel classifier visualization technique as well as demonstrating the effect of the enhancement threshold within a breast MRI CADx context.

2.1 Introduction – Classification: Dynamic Information
Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has been shown to be the most sensitive modality for screening high-risk women. Computer-aided diagnosis (CADx) systems have the potential to assist radiologists in the early detection of cancer. A key component of the development of such a CAD system is the selection of an appropriate classification function responsible for separating malignant and benign lesions. The purpose of this thesis chapter is to evaluate the use of support vector machines for the separation of malignant and benign DCE-MRI breast lesions. A classifier visualization and evaluation technique is also proposed and demonstrated.
Problems with the use and selection of the enhancement threshold (introduced in section 1.5) are also addressed. Results from the support vector machine investigation presented in this chapter were obtained with an enhancement threshold of 50%. We show that support vector machines provide an effective and flexible framework from which to base computer-aided diagnosis techniques for breast MRI, and that the proposed classifier visualization technique has potential as a mechanism for the evaluation of classification solutions. We also show that commonly used settings for the enhancement threshold can limit the CAD system’s ability to correctly label an entire lesion.

2.1.1 Methods – Classification: Dynamic Information

2.1.1.A Methods - Image Acquisition
The screening protocol used is as follows. Simultaneous bilateral magnetic resonance imaging was performed using a 1.5T magnet (GE Signa, version 11.4). Sagittal images were obtained with a phased-array coil arrangement using a dual slab interleaved bilateral imaging method [80]. This provided 3D volume data over each breast obtained with an RF spoiled gradient recalled sequence (spoiled gradient - SPGR, scan parameters: recovery time – TR / excitation time – TE / angle=18.4/4.3/30°, 256x256x32 matrix, field of view - FOV: 18x18x6-8cm). Imaging was performed before and after a bolus injection of 0.1 mmol/kg of Gd-DTPA. Each bilateral acquisition was obtained in 2 minutes and 48 seconds. Slice thickness was 2 to 3 mm.

A total of 94 DCE-MRI breast examinations from high risk patients were obtained containing lesions pathologically proven to be malignant (24 cases) or benign (70 cases). The quantity and pathological diagnosis of the different lesions addressed in this study are provided in Table III. Final diagnosis (ground truth) is based on the findings of the
histopathologist, who analyzes the tissue biopsies. Pathological diagnosis is based on a representative sample from the tissue excised from the patient. In cases where a patient with a suspicious lesion did not receive a biopsy but returned to screening for greater than one year without observed changes to the lesion, a benign diagnosis was accepted. These cases are included in Table III as having received a pathological diagnosis of “Benign by Assumption”. For the purposes of this study we have selected only those cases where the radiologist has ordered a follow-up imaging examination after attempting to diagnose the screening case. Thus we have selected a set of exams that a radiologist found difficult to separate.

**TABLE III - QUANTITY AND PATHOLOGICAL DIAGNOSIS OF BREAST LESIONS**

<table>
<thead>
<tr>
<th>Cancer/Benign</th>
<th>Pathological Diagnosis</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>Invasive Ductal Carcinoma</td>
<td>8</td>
</tr>
<tr>
<td>Cancer</td>
<td>Ductal Carcinoma in Situ (DCIS)</td>
<td>12</td>
</tr>
<tr>
<td>Cancer</td>
<td>DCIS/Microinvasion</td>
<td>3</td>
</tr>
<tr>
<td>Cancer</td>
<td>Invasive Lobular Carcinoma</td>
<td>1</td>
</tr>
<tr>
<td>Benign</td>
<td>Fibroadenoma</td>
<td>4</td>
</tr>
<tr>
<td>Benign</td>
<td>Fibrocystic Disease</td>
<td>3</td>
</tr>
<tr>
<td>Benign</td>
<td>Normal Parenchyma</td>
<td>8</td>
</tr>
<tr>
<td>Benign</td>
<td>Papilloma</td>
<td>2</td>
</tr>
<tr>
<td>Benign</td>
<td>Stromal Fibrosis</td>
<td>2</td>
</tr>
<tr>
<td>Benign</td>
<td>Ductal Hyperplasia</td>
<td>3</td>
</tr>
<tr>
<td>Benign</td>
<td>Focal Fibrosis</td>
<td>2</td>
</tr>
<tr>
<td>Benign</td>
<td>Fibrocystic Changes</td>
<td>3</td>
</tr>
<tr>
<td>Benign</td>
<td>Benign by Assumption</td>
<td>43</td>
</tr>
</tbody>
</table>

2.1.1.B *Methods - Image Registration*

Image registration is the process of aligning images that vary in position over time. This is performed in order to compensate for any patient motion that takes place during the examination. For this study we have used a 3D non-rigid registration technique for breast MRI [81] based on optical flow.
2.1.1.C Methods - Data Preprocessing and Feature Vectors

The boundary of each radiologically identified lesion was manually delineated and all of the pixels within the region of interest (ROI) were averaged together to form a single signal intensity time-series vector per lesion. The ROI was drawn around the most enhancing area of the lesion in two dimensions on the slice where the lesion is most visible in order to avoid non-lesion and necrotic tissues. The resultant curves were used in the leave-one-out validation trials presented in section 2.1.2.A. Those same regions of interest were also used to extract each signal intensity time curve within the ROI for use in the more clinically viable approach discussed in section 2.1.2.B. In this approach 5x5 median neighbourhood filtering was performed prior to extracting each time-curve in order to suppress noise. In both cases each vector has five signal intensity (SI) values corresponding to the SI obtained in the pre-contrast image and the 4 post-contrast images. We obtained a total of 94 vectors (24 cancer, 70 benign) for use in our leave-one-out classification study (section 2.1.2.A) and 2544 vectors (417 cancer and 2127 benign) for use in our randomized trials (section 2.1.2.B).

The input to any classification algorithm consists of a specific feature vector extracted from the available data. We have experimented with 4 different feature vectors related to this problem each of which is based on an individual voxel’s signal intensity (SI) time curve, where \( s(n) \) denotes the signal intensity at time \( t_n = n \times 168, n = \{0,1,2,3,4\} \) in seconds.

The four feature vectors are as follows:

1st Feature: Relative Signal Intensities. The available 5 time points are extracted and each of the terms in the resultant vector are divided by the signal intensity of the first
time point. The redundant first variable is removed and so the first feature vector consists of 4 variables and is defined as \( f_1 = \frac{s(n)}{s(0)} \) for \( n=1,\ldots,4 \). This feature was selected as it is the most straightforward approach to use.

2\(^{nd}\) Feature: Derivative of Signal Intensities. This feature vector is comprised of the time-based derivative of the relative signal intensity time curve (from feature #1) before the removal of the first redundant variable. This feature vector consists of 4 variables and is defined as

\[
\frac{d}{dt} f_1(n) = \frac{s(n+1) - s(n)}{t_{n+1} - t_n}
\]

for \( n=0,\ldots,3 \). This feature vector was selected as we hypothesize that this might reflect the actual physiological differences between malignant and benign lesions more closely. Cancerous lesions are fed by blood vessels that are characteristically leaky. It is thought that this leakiness translates into a higher rate of Gd-DTPA diffusing out of the vasculature into the lesion’s extracellular space. Since the concentration of Gd-DTPA is indirectly proportional to the observed signal intensity, it was thought that the rate of increase of signal intensity in the lesion (the derivative of the SI-time curve) might more closely reflect differences in permeability between the malignant and benign tumours. This is similar to \( K_{\text{trans}} \) (from pharmacokinetic modeling), however the derivative of the signal intensity time curve can be computed directly from the signal intensity data, whereas \( K_{\text{trans}} \) is computed through a curve fitting process that can introduce error into our permeability measurement.

3\(^{rd}\) Feature: Relative Signal Intensities and their derivatives in one vector. It was thought that the second feature may provide useful separation information, however this approach ignores the information from the first feature vector. This feature thus combines the information from the first two vectors to form a longer feature vector that consists of
8 variables and is defined as \( f_3 = \{f_1, f_2\} \). Each of the variables in this feature vector were scaled from zero to one.

4th Feature: The fourth feature vector consists of three variables: Maximum signal intensity enhancement (as a percentage) from pre-contrast to any post-contrast image, time of maximum enhancement in seconds, maximum washout (as a percentage). Each of these 3 variables were scaled from zero to one. This feature was selected as its variables are similar to the parameters used by commercially available approaches to the computer-aided diagnosis of breast MRI [48].

2.1.1.D Methods - Support Vector Machine Based Classification

Support vector machines (SVMs) are an emerging area of research in machine learning and pattern recognition [82]. SVMs are a machine learning method for creating a classification function from a set of labeled training data. Support vector machines operate by locating a decision function that attempts to split the training data into two categories. The decision function is selected such that its distance to the nearest training data on either side of the surface is maximized. If no decision function is capable of linearly separating the data, a kernel transformation function is used to map the data into a different dimensional space (called a feature space) so that it can be linearly separated using standard SVM decision function techniques. Multiple types of kernels have been developed to map data into differing dimensions. For the purpose of this study, we have compared a linear kernel (equation 1), a polynomial kernel (equation 2) and a radial basis function kernel (equation 3):

\[
K(x_i, x_j) = x_i \cdot x_j
\]  

(1)
where $x_i$ and $x_j$ are input vectors comprised of one of the previously mentioned feature vectors, $\bullet$ is the dot product operation and $\gamma$, $a$ and $d$ are kernel parameters. Non-linear kernel functions (equations 2 & 3) provide flexibility to the SVM decision function calculation. These equations allow the definition of the classification function to be non-linear in the input space. If the kernel transformation function does not fully separate the data, a slack error variable is used to create a soft margin classification function for data separation. This error variable is calculated as the weighted sum of the misclassified training set data points. For receiver operating characteristic (ROC) curve analysis performed in this thesis chapter, the relative error penalty term for our benign class was set to 1, and the error penalty term for our malignant class was varied from 1 to 7. This provides a mechanism for biasing the test towards either of our two groups, and thus provides a set of points along a ROC curve. This SVM based ROC analysis method is different from the traditional approach used in chapter 4. This non-traditional SVM based ROC analysis was selected as it biases the SVM test between the two groups in a multi-dimensional context, whereas traditional techniques are performed with a single threshold value applied to the uni-dimensional SVM regression data. ROC analysis yields a series of sensitivity and specificity values. The area under the ROC curve was computed with an absolute trapezoidal fit of these sensitivity and specificity values.
Classification of lesions has been performed as a 2-class problem where the 2 classes are cancer and benign. This classification problem has also been addressed in terms of 3-classes where the cancerous cases are separated into two classes (invasive and non-invasive cancers). For the 3-class problem, two separating classification functions were created instead of just one. The first separated invasive cancers from the third class (benign lesions). The second classification function separated non-invasive cancers from benign lesions. If either of the two classification functions classified a given lesion as cancerous, that algorithm’s prediction was considered to be cancer. Only if both decision functions classified the lesion as non-cancerous did we consider the prediction to be benign. For receiver operating characteristic (ROC) curve generation for the 3-class problem, the error penalty term for the two cancerous classes (invasive and non-invasive) were varied in unison (thus they were always identical to each other). Support vector machine based classification was implemented using the libsvm open source library [83].

2.1.1.E  Methods - Signal Enhancement Ratio Based Classification

In order to properly evaluate support vector machines as a classification mechanism for the delineation of malignant and benign lesions from DCE-MRI breast images, we compare their performance against a well established technique. We have elected to compare our approach with the commercially available signal enhancement ratio (SER) method [48]. The SER algorithm operates by calculating the signal enhancement ratio which is defined as \( \text{SER} = \frac{(s1-\text{pre})}{(s2-\text{pre})} \), where \( \text{pre} \) is the pre-contrast signal intensity. We have set \( s1 \) and \( s2 \) to be the signal intensities of our first post-contrast and last post-contrast volumes respectively. Unlike machine learning techniques, the SER algorithm does not depend on training data; the algorithm classifies a signal intensity time curve as
cancer if its SER is greater than a given cutoff value (typically set to 1.1). ROC curve
generation for the SER method was accomplished by varying the SER cutoff at which a
voxel is labelled as cancerous from 0.1 to 3.0 in steps of 0.1.

2.1.1.F Methods - Classifier Visualization / Visual Evaluation

It was thought that being able to visualize a classification function with respect to our
data set could be a beneficial means of evaluating different design options. Unfortunately,
since our data is three to eight dimensional in nature, visualization can be challenging.
Here we are presenting a new technique for visualizing a classification function that
separates data of a high dimensional nature.

The first step of this technique is to project our high dimensional data into two
dimensions. This is accomplished by a number of data projection techniques but we have
used principal components analysis. Principal components analysis is a mathematical
technique for rotating data such that the resultant orthogonal axes are aligned to the
maximum variance in the data set. This rotation is performed and the two principal
components with the highest corresponding eigenvalues are selected (referred to as the
first and second principal components). Each rotated input vector is plotted on a grid
representing these first two principal components with highest corresponding eigenvalues
(see Fig. 2.1 part A for an example plot). The second step of the technique is to sample
this principal component (PC) space across the range of first and second principal
component values (those with the highest eigenvalues) that our data set occupies. Thus
we sample from the minimum PC value in our data to the maximum PC value, however,
a larger sampling area can be used if deemed necessary. In the case of the plots presented in this thesis, principal component space was sampled 400 times along each of the first two principal components. Each of these sampled points is reverse rotated back to the input space and the resultant signal-intensity time vector is compared against any given classifier prediction technique. This generates a binary image (in our case 400x400) where the regions represent how the given classifier predicts classes in principal component space. An example binary image is provided in Fig. 2.1 part B. This binary image is processed through an edge detector (we used the Sobel edge detector – Fig. 2.1 part C) and the resultant edge is plotted in the original principal component space (where our input data has been plotted). An example plot is provided in Fig. 2.1 part D. The resultant contour line defines the border between different predictions for a given classifier in principal component space. Since we have based our projections on principal components analysis this visualization technique summarizes the total sample variance for classification interpretation. Principal components analysis was selected as it has an inverse transform (reverse rotation) which is a requirement of this proposed plotting technique, however, it should be noted that we cannot guarantee that a principal component projection is the best plane on which to evaluate a classification function. Ideally this analysis would include a projection technique that maximizes discrimination between the two classes while simultaneously supporting an inverse transform.
Fig. 2.1. An example plot of two-class data in principal component space (A), the resultant binary image (B), the edge detected image (C) and the final high dimensional classifier visualized plot in principal component space (D). These principal component space plots represent 99% of total data variance.
The percentage of total data variance displayed in the resultant two-dimensional plot can be calculated as the sum of the first two principal component eigenvalues (the two largest eigenvalues) divided by the sum of all of the eigenvalues. The classification function can also be evaluated in the lower principal components by extending this technique to a set of pair-wise plots of principal component space (plotting not just the first principal component (PC1) against the second (PC2) but also PC2 against PC3, PC3 against PC4 and so on). It should be noted that any method of projecting high-dimensional data into a two-dimensional space inherently involves the loss of some information. It is impossible to guarantee that any given projection is the ideal two-dimensional plot for classification evaluation. As such we strongly recommend sampling points in the projection space and viewing these points in the input space.

To assist in plot interpretation, input space axes were projected into the principal component space. This was accomplished by creating a fixed number (we used 100) of input space data points whereby all of the variables are set to zero but for the axis being projected. The values of the remaining variable for each input data point were scaled between 0 and a fixed percentage (we used 35%) of that variable’s total variance. This percentage value is a figure formatting parameter that controls the size of axes projected onto these two-dimensional plots. This process was repeated for each axis to be projected. The resultant input space data points were rotated into the principal component space. The resultant axes were then shifted to an appropriate viewing location. Since each axis was scaled to the same fixed percentage of data variance, the relative length of each projected axis reflects that variable’s relative influence on total data variance.
2.1.2 Results – Classification: Dynamic Information

2.1.2.A Results - Leave-One-Out Trials

Leave-one-out cross-validation consists of training the machine learning algorithm with a training set formed from all but one case of the total data set. A separating classification function was calculated based on this training set. The validation set consisted of the single remaining case and is compared with the calculated classification function. This process is repeated such that for each trial a different case is removed from the total data set for validation. This technique was conducted 94 times (once for each subject), with support vector machines attempting to classify the remaining case after training on the other 93 test cases.

The results for a linear kernel applied to each feature vector for both the 2-class and 3-class problems for leave-one-out cross-validation are provided in Table IV. The areas under each receiver operating characteristic curve are provided as per the description in the Methods (provided in Table IV under the heading Area). We have also provided the corresponding error ratio (cancer:benign under the heading ErR), sensitivity (sens), specificity (spec), positive predictive value (PPV), negative predictive value (NPV) and overall accuracy (OA) for the point on the ROC curve geometrically closest to 100% sensitivity, 100% specificity. The cancerous cutoff value for the signal enhancement ratio solution is also provided under the ErR heading. For comparative reasons we have provided the results of the SER method in each of the SVM kernel function tables (IV-VI) below.
The results for a polynomial kernel are provided in table V. The kernel parameters $\gamma$, $a$, and $d$ were varied according to the following equation: $\{\gamma, a, d\} = e^n$, where $n \in \{-7.0, 4.5\}$ in steps of 0.5. The highest areas under the ROC curve found for a polynomial kernel are provided. The table was populated in the same manner as Table IV.

### TABLE IV - RESULTS OF LEAVE-ONE-OUT TRIALS – LINEAR KERNEL

<table>
<thead>
<tr>
<th>Test</th>
<th>Class</th>
<th>Feature</th>
<th>ErR</th>
<th>Area</th>
<th>Sens. %</th>
<th>Spec. %</th>
<th>PPV %</th>
<th>NPV %</th>
<th>OA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 1</td>
<td>2</td>
<td>4.1</td>
<td>0.59</td>
<td>66.7</td>
<td>58.8</td>
<td>35.6</td>
<td>83.7</td>
<td>60.6</td>
<td></td>
</tr>
<tr>
<td>3 1</td>
<td>4.1</td>
<td>0.48</td>
<td>37.5</td>
<td>90.0</td>
<td>56.3</td>
<td>80.8</td>
<td>76.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2</td>
<td>3.1</td>
<td>0.60</td>
<td>58.3</td>
<td>64.3</td>
<td>35.9</td>
<td>81.8</td>
<td>62.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 2</td>
<td>6.1</td>
<td>0.58</td>
<td>79.2</td>
<td>37.1</td>
<td>30.2</td>
<td>83.9</td>
<td>47.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 3</td>
<td>4.1</td>
<td>0.69</td>
<td>75.0</td>
<td>52.9</td>
<td>35.3</td>
<td>86.0</td>
<td>58.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 3</td>
<td>5.1</td>
<td>0.60</td>
<td>45.8</td>
<td>75.7</td>
<td>39.3</td>
<td>80.3</td>
<td>68.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 4</td>
<td>4.1</td>
<td>0.68</td>
<td>62.5</td>
<td>71.4</td>
<td>42.9</td>
<td>84.7</td>
<td>69.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 4</td>
<td>5.1</td>
<td>0.65</td>
<td>50.0</td>
<td>78.6</td>
<td>44.4</td>
<td>82.1</td>
<td>71.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>-</td>
<td>0.69</td>
<td>62.5</td>
<td>74.3</td>
<td>45.5</td>
<td>85.2</td>
<td>71.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE V - RESULTS OF LEAVE-ONE-OUT TRIALS – POLYNOMIAL KERNEL

<table>
<thead>
<tr>
<th>Test</th>
<th>Class</th>
<th>Feature</th>
<th>$\gamma$</th>
<th>$a$</th>
<th>$d$</th>
<th>ErR</th>
<th>Area</th>
<th>Sens. %</th>
<th>Spec. %</th>
<th>PPV %</th>
<th>NPV %</th>
<th>OA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 1</td>
<td>2</td>
<td>2.72</td>
<td>2.72</td>
<td>1.65</td>
<td>4.1</td>
<td>0.66</td>
<td>75.0</td>
<td>55.7</td>
<td>36.7</td>
<td>86.7</td>
<td>60.6</td>
<td></td>
</tr>
<tr>
<td>3 1</td>
<td>0.05</td>
<td>0.37</td>
<td>2.72</td>
<td>4.1</td>
<td>0.65</td>
<td>33.3</td>
<td>97.1</td>
<td>80.0</td>
<td>81.0</td>
<td>80.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2</td>
<td>0.37</td>
<td>2.72</td>
<td>2.72</td>
<td>3.1</td>
<td>0.65</td>
<td>66.7</td>
<td>65.7</td>
<td>40.0</td>
<td>85.2</td>
<td>66.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 2</td>
<td>1.00</td>
<td>2.72</td>
<td>2.72</td>
<td>5.1</td>
<td>0.67</td>
<td>70.8</td>
<td>64.3</td>
<td>40.5</td>
<td>86.5</td>
<td>66.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 3</td>
<td>2.72</td>
<td>1.65</td>
<td>2.72</td>
<td>3.1</td>
<td>0.70</td>
<td>66.7</td>
<td>65.7</td>
<td>40.0</td>
<td>85.2</td>
<td>66.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 3</td>
<td>0.37</td>
<td>0.001</td>
<td>1.00</td>
<td>5.1</td>
<td>0.66</td>
<td>41.7</td>
<td>90.0</td>
<td>58.8</td>
<td>81.8</td>
<td>77.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 4</td>
<td>0.22</td>
<td>0.14</td>
<td>2.72</td>
<td>3.1</td>
<td>0.71</td>
<td>62.5</td>
<td>78.6</td>
<td>50.0</td>
<td>85.9</td>
<td>74.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 4</td>
<td>0.03</td>
<td>2.72</td>
<td>2.72</td>
<td>5.1</td>
<td>0.71</td>
<td>62.5</td>
<td>78.6</td>
<td>50.0</td>
<td>85.9</td>
<td>74.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.9</td>
<td>0.69</td>
<td>62.5</td>
<td>74.3</td>
<td>45.5</td>
<td>85.2</td>
<td>71.3</td>
<td></td>
</tr>
</tbody>
</table>
The results for a radial basis function kernel are provided in Table VI. The kernel parameter $\gamma$ was varied according to the following equation: $\gamma = e^n$, where $n \in \{-7, 0, 4.5\}$ in steps of 0.1. The highest areas under the ROC curve found for a radial basis function kernel are provided. The table was populated in the same manner as Tables IV and V.

**TABLE VI - RESULTS OF LEAVE-ONE-OUT TRIALS – RADIAL BASIS FUNCTION KERNEL**

<table>
<thead>
<tr>
<th>Class</th>
<th>Feature</th>
<th>$\gamma$</th>
<th>ErR</th>
<th>Area</th>
<th>Sens.</th>
<th>Spec.</th>
<th>PPV</th>
<th>NPV</th>
<th>OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0.30</td>
<td>4:1</td>
<td>0.63</td>
<td>75.0</td>
<td>37.1</td>
<td>29.0</td>
<td>81.3</td>
<td>46.8</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.07</td>
<td>5:1</td>
<td>0.63</td>
<td>33.3</td>
<td>91.4</td>
<td>57.1</td>
<td>80.0</td>
<td>76.6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.15</td>
<td>3:1</td>
<td>0.64</td>
<td>45.8</td>
<td>81.4</td>
<td>45.8</td>
<td>81.4</td>
<td>72.3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4.06</td>
<td>5:1</td>
<td>0.60</td>
<td>54.2</td>
<td>64.3</td>
<td>34.2</td>
<td>80.4</td>
<td>61.7</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2.46</td>
<td>3:1</td>
<td>0.68</td>
<td>62.5</td>
<td>71.4</td>
<td>42.9</td>
<td>84.7</td>
<td>69.1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.20</td>
<td>5:1</td>
<td>0.64</td>
<td>37.5</td>
<td>88.6</td>
<td>52.9</td>
<td>80.5</td>
<td>75.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td><strong>4.06</strong></td>
<td>3:1</td>
<td><strong>0.74</strong></td>
<td><strong>62.5</strong></td>
<td><strong>78.6</strong></td>
<td><strong>50.0</strong></td>
<td><strong>85.9</strong></td>
<td><strong>74.5</strong></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.09</td>
<td>5:1</td>
<td>0.71</td>
<td>62.5</td>
<td>78.6</td>
<td>50.0</td>
<td>85.9</td>
<td>74.5</td>
</tr>
<tr>
<td>SER</td>
<td>-</td>
<td>-</td>
<td>0.9</td>
<td>0.69</td>
<td>62.5</td>
<td>74.3</td>
<td>45.5</td>
<td>85.2</td>
<td>71.3</td>
</tr>
</tbody>
</table>

The highest area under the ROC curve was obtained when using a radial basis function kernel with $\gamma$ set to 4.06 on the fourth feature vector for the 2-class problem. The area under the curve was 0.74 (corresponding entry highlighted in Table VI).

We implemented the plotting technique described in section 2.1.1 on the radial basis function kernel classifier that maximized the area under the ROC curve (feature vector 4, $\gamma=4.06$, ROC Area=0.74, see bolded entry in Table VI) for the leave-one-out trials (94 vectors). We fixed the error ratio at 3:1 (cancer:benign) as this produced the best ROC area values (see Table VI). The resultant plot is provided in Fig. 2.2.
Fig. 2.2. Contour line (blue) of a radial basis function kernel classifier ($\gamma = 4.06$, error ratio 3:1) in principal component space (94% of total data variance displayed from feature vector 4) – bolded entry from Table VI.

### 2.1.2.B Results - Pixel-by-Pixel Randomized Validation

Although leave-one-out validation is a useful tool for evaluating a given approach, we were also interested in the robustness of the algorithms. It is also beneficial to evaluate these approaches in a more clinically viable manner. For the purposes of this study we did not differentiate between benign lesions and normal tissue (both are simply considered non-cancer). All support vector machine results were obtained using the third feature vector (relative signal intensities and their derivatives) and the fourth feature vector (Maximum signal intensity enhancement, time to maximum enhancement, maximum washout). The first and second feature vectors were not considered in this section as they consistently underperformed the third and fourth feature vectors in the previous leave-one-out trials.
We have conducted a series of 100 random trials whereby 75% of the cases (94 in all) from each class were randomly selected and used for training (2544 vectors in all, obtained from the most intense regions of our 94 lesions). The calculated classification function was then applied to the remaining examinations (validation set) on a pixel-by-pixel basis across the whole parenchyma tissue on the slice where the lesion is most visible. If at least one pixel within a radiologically identified cancerous lesion was labelled as cancer the case was counted as a true positive. If any pixel in a radiologically identified benign lesion or normal breast parenchyma was classified as cancer, the case was counted as a false positive. This enumeration scheme was employed as we know the histopathologist’s diagnosis of each radiologically identified lesion, but we have no official gold standard diagnosis for each voxel. Noise removal was performed by 5x5 neighbourhood median filtering. The results for the signal enhancement ratio (SER) method were included as well and have been subjected to the same noise filtering and test evaluation metrics as the support vector machine approaches. All classification methods were subjected to a 50% enhancement threshold, whereby voxels whose maximum enhancement from the pre-contrast volume to any post-contrast volume is less than the given threshold are classified as non-cancerous. This threshold was selected as it is in-line with thresholds quoted in the literature [41, 48] and was determined to be below the minimum enhancement of all cancerous lesions in our data set (where the calculated enhancement is averaged across all the hyperintense voxels in the given lesion). The effects of varying the enhancement threshold in CAD breast MRI are demonstrated below.
As was the case in the leave-one-out trials, ROC curve generation was accomplished for both methods (SVM and SER) by the same techniques outlined in the sections 2.1.1.D and 2.1.1.E. The resultant test parameter means and standard deviations (in brackets) for the test case on the ROC curve whose sensitivity and specificity values are geometrically closest to 100%, 100% are provided in Table VII for features 3 and 4. For each SVM test case, any relevant kernel parameter and error ratio settings are provided. In an effort to thoroughly evaluate the classification functions, we have also tracked the number of false positive pixels per cancerous examination (FPP) and the number of true positive pixels per cancerous examination (TPP). Since the SER method does not require training, it was applied on a pixel-by-pixel basis on all 94 examinations.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Test Kernel</th>
<th>y, A, d</th>
<th>EPR</th>
<th>Sens. %</th>
<th>Spec. %</th>
<th>PPV %</th>
<th>NPV %</th>
<th>OA %</th>
<th>FPP</th>
<th>TPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Linear</td>
<td>-</td>
<td>4.1</td>
<td>0.54 (0.13)</td>
<td>55.5 (21.7)</td>
<td>48.3 (24.1)</td>
<td>77.2 (12.9)</td>
<td>73.8 (14.6)</td>
<td>50.1 (5.8)</td>
<td>16.3 (35.5)</td>
</tr>
<tr>
<td>4</td>
<td>Linear</td>
<td>-</td>
<td>3.1</td>
<td>0.40 (0.09)</td>
<td>62.0 (32.4)</td>
<td>35.8 (35.9)</td>
<td>69.8 (9.8)</td>
<td>60.7 (24.6)</td>
<td>40.1 (20.1)</td>
<td>36.5 (41.2)</td>
</tr>
<tr>
<td>3</td>
<td>Poly 0.37, 0.61, 2.72</td>
<td>4.1</td>
<td>0.62 (0.1)</td>
<td>44.5 (22.6)</td>
<td>73.4 (25.0)</td>
<td>43.1 (22.6)</td>
<td>76.4 (15.4)</td>
<td>60.2 (15.2)</td>
<td>17.7 (59.7)</td>
<td>26.0 (27.5)</td>
</tr>
<tr>
<td>4</td>
<td>Poly 2.72, 10.272</td>
<td>2.1</td>
<td>0.59 (0.1)</td>
<td>41.7 (9.6)</td>
<td>76.3 (16.2)</td>
<td>51.6 (21.3)</td>
<td>79.6 (5.5)</td>
<td>67.7 (21.2)</td>
<td>5.6 (10.0)</td>
<td>25.6 (24.1)</td>
</tr>
<tr>
<td>3</td>
<td>RBF 0.22</td>
<td>3.1</td>
<td>0.62 (0.11)</td>
<td>39.8 (18.3)</td>
<td>85.1 (10.8)</td>
<td>48.9 (23.2)</td>
<td>81.1 (4.7)</td>
<td>73.7 (4.2)</td>
<td>5.1 (10.9)</td>
<td>20.5 (20.5)</td>
</tr>
<tr>
<td>3</td>
<td>RBF 0.22</td>
<td>4.1</td>
<td>0.62 (0.11)</td>
<td>55.0 (21.9)</td>
<td>60.1 (20.9)</td>
<td>36.5 (17.3)</td>
<td>76.2 (16.4)</td>
<td>58.8 (10.8)</td>
<td>18.4 (44.7)</td>
<td>35 (50.1)</td>
</tr>
<tr>
<td>4</td>
<td>RBF 1.85</td>
<td>2.1</td>
<td>0.57 (0.1)</td>
<td>42.2 (8.7)</td>
<td>73.2 (19.5)</td>
<td>38.8 (22.4)</td>
<td>78.7 (7.3)</td>
<td>65.5 (2.6)</td>
<td>6.3 (11.2)</td>
<td>25.3 (22.9)</td>
</tr>
<tr>
<td>SER</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>0.49</td>
<td>58.3</td>
<td>50.0</td>
<td>28.6</td>
<td>77.0</td>
<td>52.1</td>
<td>34.5</td>
</tr>
</tbody>
</table>

We implemented the plotting technique (section 2.1.1.F) on two of our radial basis function classifiers from the randomized trials (2544 vectors) in Fig. 2.3 for feature vector 3. Contour lines are provided for SVM solutions with error penalty ratios of both 3:1 and 4:1. The classifier contour line for the SER approach is provided along with the
contour line representative of the 50% enhancement threshold. The test parameters corresponding to the two SVM classifiers and the SER classifier are bolded in Table VII. In order to assist in the interpretation of Fig. 2.3, we have sampled four points in principal component space (seen on Fig. 2.3 as black diamonds at locations (1,0.7), (0.7,0), (0,0.4), and (0,0)) and the corresponding signal intensity time curves are provided in Fig. 2.4.

**Fig. 2.3.** Contour lines (blue) of a radial basis function kernel classifier \((\gamma=0.22,\) error ratio 3:1 – dashed line, error ratio 4:1 – dotted line) and the SER based classifier (cancer cut off 1.1, solid magenta line) in principal components space (88% of data variance displayed from feature vector 3).
**Fig. 2.4.** Signal intensity time curves of points sampled in principal component space (black diamonds from Fig. 2.3).

We were interested in comparing a scatter plot showing both invasive and non-invasive cancers to help shed light on the results of our 3-class problem experiments and provided the plot in Fig. 2.5 (plotted from feature vector 3).

**Fig. 2.5.** Invasive and non-invasive cancers plotted in principal component space (91% of total data variance displayed from feature vector 3).
We have demonstrated how the results of support vector machine classification can be visualized for radiological analysis in Fig. 2.6. Here we have colour coded any pixel that fell on the malignant side of the support vector machine classification function in red. The automated support vector machine based diagnosis of a single 256x256 slice takes about 3 seconds on a 3 GHz Intel Pentium 4 workstation.

**Fig. 2.6.** A ductal carcinoma in situ (left) correctly labelled as cancer by a radial basis function ($\gamma=0.22$, error ratio 3:1) support vector machine based classifier (right).

Fig. 2.7 is provided to demonstrate how increases in the enhancement threshold can compromise a test’s ability to correctly classify the full size of a lesion. Here pixels labelled as cancerous have been colour coded red by the CADx method addressed. This is significant as a reduction in the size of a CADx detected lesion may influence a radiologist’s decision on a patient’s care, and could result in a missed cancer. This can occur if the radiologist relies on the CADx results but dismisses very small clusters of potentially malignant tissues due to the false positive rate. It can be seen in Fig. 2.7 that at a threshold of just 75%, much of the lesion has been classified as non-cancerous. It is
clear that some of the less enhancing areas of our cancerous lesions are adversely affected by low thresholds. Fig. 2.7 shows that thresholds can limit a CADx test’s ability to correctly identify the full size of a lesion and as such should not be raised above 60%. This is notable with respect to numerous prior studies that have made use of higher enhancement thresholds [42, 62-72].

**Fig. 2.7.** A sagittal image with an invasive ductal carcinoma (A – see arrow) magnified and labelled by the SVM method at 0% (B), 75% (C) and 100% (D) thresholds. The images were acquired using the MRI protocol described in the Methods.
These examinations were also analyzed by a trained radiologist blinded to the findings of any of the computer-aided diagnosis techniques. The radiologist diagnosed each examination on the BI-RADS scale [31]. The examinations considered in this study were also deemed difficult to diagnose by the radiologist. Of the total 94 examinations, 76 were rated BI-RADS 0 (Needs further work-up). Of the remaining 18 cases there were 5 true positives, 5 false positives, 8 true negatives and 0 false negatives when we consider BI-RADS 4 and 5 as a cancerous diagnosis and BI-RADS 1-3 as benign.

We compared the ROC area distributions for feature 4 (polynomial kernel) and feature 3 (RBF kernel) and the calculated probability that the observed differences (bolded entries in Table VII) occurred randomly was 0.0012. This indicates that the observed differences between feature 3 and feature 4 were significant. We also ran the same statistical significance test comparing feature 3 (polynomial kernel) with feature 3 (RBF kernel) and found that the probability that the observed differences occurred randomly was 0.6762, indicating that these two are not statistically significant.

2.1.3 Discussion – Classification: Dynamic Information

We have evaluated support vector machines as a potential mechanism for the design of a classifier responsible for delineating between malignant and benign breast lesions from DCE-MRI time-series data. There were a number of motivations for selecting SVMs as a classification mechanism. SVMs have been shown to perform well in medical applications such as computer-aided diagnosis [36], and have also been shown to perform well even with relatively small training sets [34]. This was particularly appealing given the inherent difficulty in acquiring large amounts of screening data devoted exclusively to
training. Support vector machines also perform well and classify reasonably quickly on high dimensional data. We evaluated SVMs by comparing their performance with the signal enhancement ratio (SER) based breast MRI computer-aided diagnosis method. We selected the SER method as it is both commercially available (Confirma Inc., Kirkland WA USA) and currently in use by radiologists.

We provided a visual plot of the classification function with the highest ROC area for leave-one-out trials in Fig. 2.2. For feature vector 4, our data was divided into four clusters that reflect the fourth feature’s second variable (the time in seconds to peak enhancement). Since there were only four post-contrast volumes acquired there are only four possible values for this time to enhancement term. We can see from this plot that the classification function is simply dividing the data based on its time to enhancement. This was easily visualized with the provided axis projections and demonstrates the benefit of the proposed classifier visualization technique as a method to effectively interpret the behaviour of a classification function.

Although we have provided results for the cases with the highest receiver operating characteristic (ROC) curve areas, support vector machines produced many computer-aided diagnostic design alternatives (examples provided in Table VII). It should be noted that the standard deviations in Table VII are relatively high. It was most likely due to insufficient tuning of the error penalty term to suit the training data for a given trial. High standard deviations could also be attributed to the inseparability of our data due to both the patient population and the temporal resolution of our MRI acquisition protocol. Since
our data was not normally distributed we have elected to test for statistical significance using the Wilcoxon-Signed Rank Test which shows a statistically significant difference between feature 3 and feature 4.

The signal enhancement ratio method is based on a linear equation, consequently the classifying border in principal component space is also linear. It should be noted that the observed nonlinear regions of the signal enhancement ratio boundary function (see Fig. 2.3) are in fact caused by the 50% threshold (border provided in black) we applied to all classifiers. Future research will experiment with varying this threshold over a wide range of values to determine its effect on the operation of classification based computer-aided diagnostic systems. It can be seen in Fig. 2.3 that the 50% threshold border divides the inseparable data forcing a non-cancerous diagnosis of a few cancerous vectors. It is thought that despite efforts to draw the regions of interest around the most enhancing part of each lesion there was still some variance within some of our cancerous tumours.

Although feature 3 resulted in the best performance (see Table VII), it was compromised by being fixed to the MRI acquisition protocol (ie. if the protocol were changed, the measurement will no longer be valid). In theory feature 4 should be able to function on any acquisition protocol, and so although it performs somewhat worse than feature 3, it remains an interesting design option. Future work will involve evaluating feature 4 on a variety of MRI acquisition protocols.
It is difficult to perform a direct comparison between the results of this study and those obtained by the radiologist. This data set was particularly challenging to separate. In fact the radiologist diagnosed 81% of the exams as BI-RADS 0 (needs a more detailed imaging examination). Since the CADx techniques presented do not label lesions as BI-RADS 0, a proper comparison is not possible.

We experimented with dividing our group of cancers into two classes (invasive and non-invasive cancer) and referred to these experiments as a 3-class problem (the third class is benign). This division was made as invasive cancer is characterized by having spread into neighbouring tissues and through angiogenesis, the tumours are supplied with characteristically leaky blood vessels. This increased permeability increases the tumour tissue’s uptake and washout rates of our Gadolinium based (Gd-DTPA) contrast agent. Thus we expect to see different signal-intensity time curves for invasive cancer as compared to non-invasive cases which have not spread into neighbouring tissues. In practice the 3-class problem consistently underperformed the 2-class approach for all kernel functions and as such we elected to forgo the 3-class problem for the randomized trials presented in section 2.1.2.B. Ultimately the efficacy of solving the problem with two independent classification functions will depend on how distinct the two classes (invasive and non-invasive) are. As can be seen in Fig. 2.5, although there is some separation between invasive and non-invasive cancers, the two groups overlap each other significantly. These conclusions are supported by recent research which shows that the dynamic characteristics of non-invasive lesions is variable [84] as they contain all of the range of expected signal intensity curve types (washout, plateau and consistent
enhancement). This indicates that perhaps creating two classification functions (one for each invasive and non-invasive cancers) will not assist in the delineation between malignant and benign tissues.

When considering feature vectors, the first (relative signal intensities) was selected as it is the most obvious choice. As mentioned in section 2.1.1.C features 2 and 3 were selected due to the relationship between the derivative of the signal intensity time curve and the permeability of the vasculature in cancerous tumours. It was the third feature vector (both relative signal intensities and their derivatives) that performed best in the more clinically realistic pixel-by-pixel trials (probably due to the inclusion of all the relevant feature data). The additional information in the third feature vector (relative SI and their derivative) is present in the original data set (feature #1 – relative signal intensities) in the form of the relationship between adjacent data points. However, it is believed that including these derivatives explicitly in the feature vector facilitates the exploitation of these differences by support vector machines. Ideally, we would like to convert the signal intensity time curves into concentration time curves since the concentration does not depend on the magnetic resonance imaging sequence used. This, however, requires a pre-contrast map of $T_1$ values which was not available in this study. Our current classification algorithm uses the relative signal intensity values from the MRI screening images as 3 to 8 dimensional “feature vectors”. If we were to apply feature vectors 1-3 with a higher temporal resolution, the size of the feature vector would increase, however it is known that the stability of a classification algorithm generally decreases as the number of dimensions increases [85], so caution is warranted. Scaling was performed on our feature
vectors so that the individual features within each vector are scaled to occupy the same range. Scaling is performed so that one feature with an initially larger scale does not dominate over the other features during classification. One way of overcoming this problem is to parameterize the signal intensity time curves using pharmacokinetic modeling [86]. It has been mentioned that pharmacokinetic modeling involves a curve fitting phase that is subject to error. If the number of time points acquired in the MRI acquisition protocol were increased, then the pharmacokinetic fit would become more reliable. Since our temporal acquisition protocol was too coarse for pharmacokinetic modeling, a simple empirical curve estimation method was used in feature #4. Furthermore improvements to the acquired image’s signal to noise ratio will also allow for more reliable pharmacokinetic fitting.

Analysis of the lesions in this data set indicated that raising the enhancement threshold above 100% would result in an inability to segment some of the cancerous lesions [79]. This effect is unacceptable in the context of a computer-aided detection technique and so the enhancement threshold was not raised over 100%. Additionally an enhancement threshold below 40% consistently grouped large amounts of non-lesion tissue with the desired lesion making threshold settings this low unacceptable [79]. Additionally, it should be noted that none of the studies making use of the enhancement threshold used values lower than 40% [41-42, 48, 62-73].

Ideally, any future computer aided diagnostic system will have the flexibility to accommodate a variety of magnetic resonance imaging protocols without compromising
the separability of the data. In addition to the dynamic information, spatial patterns have also been shown to provide considerable discrimination between malignant and benign breast MRI lesions [52]. Chapter 3 will examine techniques to extract spatial features in addition to the dynamic information already considered in this study. Additionally, in order to form firm conclusions about the efficacy of the discussed methods, validation will need to be performed on independent data sets.
Chapter 3

Segmentation

Some of the material presented in this thesis chapter is the subject of a successful grant application from the Ontario Research and Commercialization Program Grant and another successful grant application from the Medical Technologies Research and Commercialization Collaborative. Furthermore, a Canadian Breast Cancer Foundation fellowship award was received to (in part) develop the work presented in this chapter.
This thesis chapter is concerned with region-of-interest identification (segmentation) towards the detection of suspicious lesions from breast MRI examinations that may represent malignancies. In this chapter, the segmentation phase of a breast MRI CAD system is addressed exclusively. Segmentation techniques which combine pixels within an examination to be part of the same region-of-interest are evaluated and a new technique is presented. This thesis chapter is concerned with comparatively evaluating the proposed segmentation technique with the most widely used segmentation method for breast MRI examinations.

3.1 Introduction - Segmentation

Image analysis is performed in a diverse set of research fields with widely varying applications such as remote sensing / satellite imaging, robotics, computer vision, security applications and medical imaging. A common first step in image interpretation is the process of identifying regions-of-interest, a process referred to in the literature as image segmentation or clustering. Many methods have been developed to perform this task, however, most existing techniques have considerable limitations, particularly in their need to be fine tuned to the specific application at hand. The purpose of this thesis chapter is to present a robust method for the identification of regions-of-interest in varying image types and to comparatively evaluate the technique in a breast MRI CADe context.
3.1.1 Methods – Segmentation

3.1.1.A Methods – Segmentation – A Hierarchical Approach

Region-of-interest (ROI) identification (segmentation/clustering) takes an image as input and outputs an assigned group value for each pixel (dot) in the image. This can be thought of as a basic amount of image interpretation whereby a set of pixels with the same group number are said to represent a single region-of-interest. Our proposed ROI identification technique is a hierarchical agglomerative bottom-up clustering/segmentation method whereby all pixels in the image are initialized to unique group values. The hierarchical nature of this algorithm is illustrated in Fig. 3.1 whereby an example input image is provided (upper left) and is subdivided into small panes each of which consists of a preset number of pixels (as marked by the green lines – upper centre). Each of these groups were initialized to unique group values. Each set of 4 panes (2x2) is segmented by the same algorithm (called a node and represented by a blue square – upper right). This 2x2 merger process is repeated hierarchically causing the panes to become progressively larger (lower left) until the final node merges 4 panes that constitute the entire image (lower centre). An example final segmentation is provided (lower right).
Fig. 3.1. A demonstration of the hierarchical agglomerative approach used. An example input image is provided (upper left). The image is divided into panes as marked by green lines (upper centre). Four adjacent panes are combined by a node represented by a blue square (upper right) forming a merged segmentation. This is repeated hierarchically causing the panes to be progressively larger (lower left) until the final node merges 4 panes that constitute the entire image (lower centre). An example final segmentation is provided (lower right).


The merging process takes place along the borders of each input pane and simply needs to decide if the two groups on either side of the pane border are part of the same group or are significantly different. This node merger process is illustrated in Fig. 3.2 whereby the upper half of the figure demonstrates a node initially performing a merger of its upper and lower panes (mergers are taking place along the green line that borders the two panes). When looking at the left half of the image being merged about the green line, groups 1&5, 2&6 and 2&5 will be compared for a potential merger. Groups 1&6 will not
be compared directly as they do not share a border. Note that groups 2&5 will not be

![Diagram of pane merger process]

**Fig. 3.2.** An overview of the recurring pane merger process. First the vertically connected panes are merged about a horizontal line (marked in green in the upper right). The two resulting panes are merged about a vertical line (lower half). The output image pane is provided in the bottom right.

merged but the other pairings will, leading to the segmentation shown in the lower left of the figure. This top and bottom merger process represented in the upper half of Fig. 3.2 is followed by the merger of the resulting left and right panes as is illustrated in the lower half of Fig. 3.2. Only groups which share a border along these pane merger lines are compared for potential merging. The decision as to whether or not two groups are the same is accomplished through statistical testing (metric provided below in equation 4). When statistics cannot be reliably performed (for instance when not enough samples are
available in the two groups), a simple Euclidean distance measure between the means of the pixel values of the two groups is used to determine whether the groups are the same or different. The statistical merger cutoff is this program’s predominant input parameter and acts as a threshold on equation 4 below. For most applications this should be the only parameter that needs to be set by the user and for many applications it should not need to be changed from its default setting of 0.6. Setting this parameter to different values has the effect of creating regions-of-interest at different physical sizes. Appropriate settings of the parameter can be tuned by the researcher.

3.1.1.C Methods – Segmentation – Border Refinement

After a node’s merger process was completed any ROI borders that exist along the pane merger lines are subjected to an optional refinement step (all of the results in this thesis were generated with this border refinement enabled). In this stage neighbouring groups with enough samples for reliable statistics had their borders modified so as to minimize the resultant separability metric (Eq. 4 below) calculated between the two groups. This has the effect of maximizing the separation between the two groups. This is accomplished by considering a border between groups to be a threshold (like a contour line on a topographic map) and evaluating a range of possible threshold borders between the two groups. For each possible border the resultant two groups are compared using a custom metric (equation 4) and the case with the most separable measurement is selected as the final border.
3.1.1.D Methods – Segmentation – Statistical Comparisons

We have experimented with a variety of statistical techniques that input two groups of numbers and output a single measurement of how separable the two groups are (2 sample t-test, Mann-Whitney U test, etc.). However, existing statistical significance tests involve assumptions regarding the underlying data, for instance the t-test assumes the data is bell-shaped (Gaussian distribution) and the Mann-Whitney U test assumes the data is symmetric about the median. To avoid making such assumptions we found it useful to create a new statistical sorting metric in order to achieve our desired results. The test measures the distance between two groups of numbers, then measures the maximum distance these two groups of numbers could have obtained if they were completely separable. This is achieved by sorting the joint distribution of the two groups and assigning the lowest sorted samples to the lower group and the highest samples to the upper group and then measuring the distance between the two artificially separated groups. This metric is defined in equation 4. Measuring the distance between the two groups can be accomplished in many ways, including calculating the sum of the distances between the members of each group, or simply measuring the distance between two standard location measures such as the mean or median (the results in this thesis were generated with the mean distance measure). The user selects a threshold value for this D metric which acts as a decision function regarding whether or not to merge the two groups.
\[ D = 1 - \frac{|X - \bar{Y}|}{nZ - Z''} \]  \hspace{1cm} (4)

Where, 
- \( X \) is the lower valued \((X < \bar{Y})\) input distribution with \( n \) samples
- \( Y \) is the upper valued \((X < \bar{Y})\) input distribution with \( m \) samples
- \( Z \) is the sorted joint distribution \( \{X;Y\} \)
- \( Z'' \) is the terms of \( X \) among the lowest \( n \) terms of the joint distribution \( Z \)
- \( Z''' \) is the terms of \( Y \) among the highest \( m \) terms of the joint distribution \( Z \)
- \( n \) is the number of samples in distribution \( X \)
- \( m \) is the number of samples in distribution \( Y \)

The aforementioned statistical techniques for comparing two groups of numbers are methods whereby there is a single sample per pixel (for example a grayscale image). When there are more samples per pixel, such as in DCE-MRI, then standard statistical techniques cannot be used as they only account for a single measurement. This problem can be overcome by reducing our set of multiple measurements to a single measurement per sample. The color image results generated in this thesis section were obtained by projecting multi-sample data into a single-sample space. This projection can be accomplished by a variety of methods, for example, using principal components analysis or custom projection techniques. The results provided in this thesis were obtained by projecting the data from the two groups being compared onto the axis that connects the means of those two groups. This multidimensional to unidimensional projection step is performed as a preliminary step in statistical testing and is illustrated in Fig. 3.3 below. The example illustration assumes that three measurements have been made per sample (although any number of measurements is possible). Example data plotted against these three measurements are provided in the upper left pane. In the upper right pane the mean values of each group are provided as crosses. In the lower left pane a green axis is
provided that connects the mean values of the two groups. Finally, in the lower right pane the data is shown projected onto this green axis.

![Fig. 3.3. Example data with multiple measurements per subject (upper left), with means provided as crosses (upper right), with a green axis connecting the two means (lower left), and the data projected onto the green axis (lower right).](image)

3.1.1.E Methods – Segmentation – Breast MRI

The purpose of this section is to present an adaptation of the aforementioned automatic multi-spectral segmentation algorithm that supports the segmentation of volumetric...
breast MRI examinations. This section also comparatively evaluates the proposed segmentation technique against the most common technique for segmenting lesions from magnetic resonance breast images that supports automatic region-of-interest selection: the enhancement threshold. The traditional approach for evaluating segmentation techniques is to compare each method against a gold-standard, however, no definitive gold-standard exists for breast MRI. The best that could be done would be to ask one, or preferably more, expert radiologists to perform manual segmentations. This is very time consuming and therefore is usually only carried out on a small subset of studies. There is also the problem of inter-observer variability and, particularly in the case of very small lesions, it is not clear that a manual segmentation is a satisfactory gold-standard. To circumvent these shortcomings, evaluation in this thesis chapter was performed within a computer-aided detection context, that is, the ability of each segmentation approach to support the extraction of feature measurements that assist in the discrimination between malignant and benign lesions is assessed. This atypical evaluation methodology is illustrated in Fig. 3.4, whereby we are comparing the amount of relative separation between malignant and benign lesions (measured by ROC area) as a function of the segmentation algorithm’s input parameter.

Fig. 3.4. Block diagram representing the atypical segmentation evaluation methodology used in this thesis chapter.
3.1.1.F Methods – Segmentation – Image Acquisition

The screening protocol used is identical to that used in the thesis sections 2.1.1, however, in this section an expanded data set of 223 DCE-MRI breast examinations was used (44 malignant and 179 benign).

3.1.1.G Methods - Segmentation – Image Registration

Image registration is the process of aligning images that vary in position over time. This is performed in order to compensate for any patient motion that takes place during the examination. For this thesis section (as in all the others) we have used a 3D non-rigid registration technique for magnetic resonance breast images [81].

3.1.1.H Methods – Proposed Volumetric Segmentation

The main extension required to enhance this technique towards volumetric images (or video sequences) is for a single node to input 8 previously segmented cubes instead of 4 previously segmented panes. The proposed segmentation method begins by dividing the entire examination into volumetric blocks each containing 4x4x2 pixels/voxels. Each volumetric block is initialized to a unique group number. Although this initially assigns a contiguous set of 32 pixels to being part of the same group, the technique is able to detect smaller sized lesions by means of the edge refinement step previously described. The full set of volumetric blocks that constitute the entire examination are agglomerated in a traditional hierarchical pattern. The agglomeration process uses a discriminating function to determine whether two neighbouring groups are part of the same group or represent two different groups. The discriminating function measures the amount of mutual overlap.
observed between the two distributions and was provided in equation 4. For the purposes of this study, two neighbouring groups are merged if the computed metric from equation 4 is greater than a threshold value. This threshold value was varied from 0.55 to 0.65 in steps of 0.01.

### 3.1.1.1 Methods – Comparative Segmentation

In order to evaluate our proposed segmentation technique, we must compare the results of our technique with those obtained by an established segmentation method. The most common technique that supports the automatic segmentation of entire breast MRI examinations is the enhancement threshold [41-42, 48, 62-73]. An MRI breast examination measures the signal intensity throughout the breast. The examination then involves the injection of a contrast agent followed by more signal intensity measurements. Malignant tissue tends to be fed by characteristically leaky blood vessels which allow delivery of the contrast agent. Suspicious tissues exhibit enhancement over time. We compute the maximum amount of enhancement observed at each voxel location in equation 5. A previous study exclusively investigated the effect of the enhancement threshold in computer-aided diagnosis systems [79] and showed that the selection of the enhancement threshold affects the ability of CADe systems to correctly segment small cancerous lesions.

\[
E_{\text{max}} = \frac{\max \left( \left\{ SI_i \ldots SI_{FTP} \right\} \right)}{SI_0}
\]

Where, \( SI_i \) is the signal intensity at time point \( i \)

\( FTP \) is the final time point
All of the voxels that exceed the provided enhancement threshold are converted into independent contiguous regions-of-interest for the feature measurements below. For the purposes of this study, enhancement threshold based segmentation was accomplished by varying the threshold from 40% to 100% in steps of 5%.

3.1.1 Methods – Segmentation – Feature Measurements

In order to evaluate how much separation can be obtained between malignant and benign lesions, a series of measurements (often called features) were selected. Each measurement listed below was calculated for all 223 lesions segmented by both techniques at all relevant parameter settings as outlined in sections C and D above.

Feature Measurement #1: Average Time Normalized Enhancement

The time normalized enhancement measurement for a single voxel is defined in equation 6. The technique computes the enhancement as specified in equation 5 for each voxel in our suspect lesion and normalizes by the time to peak in seconds. The average of these time normalized enhancement values form the final lesion measurement. The measured enhancement and time to peak enhancement are directly linked to the permeability of the vasculature in the lesion.

\[ T_{NE} = \frac{E_{max}}{T_{E_{max}}} \] (6)

Where, \( T_{E_{max}} \) is the time to maximum enhancement in seconds.
**Feature Measurement #2: Average Enhancement Reduction**

A breast MRI examination involves the injection of a contrast agent which will pool in lesion tissue. Over the course of a breast MRI examination the concentration of contrast agent in the blood plasma will diminish as contrast agents that have leaked into a lesion may begin to diffuse out of the lesion and back into the blood plasma – a process commonly referred to as washout. Enhancement reduction measurements are an alternative measure of lesion vascularity. A single voxel’s enhancement reduction measurement is provided in equation 7 and is measured as the percentage of decreased signal intensity with respect to the maximum signal intensity enhancement measured in equation 5. An enhancement reduction value is calculated for each voxel in our region-of-interest (ROI) and the average of these values form the average enhancement reduction feature measurement.

\[
E_R = 1 - \frac{E_{\text{min}}}{E_{\text{max}}}
\]  

(7)

Where, \(E_{\text{min}}\) is the minimum enhancement after \(T_{\text{max}}\).

**Feature Measurement #3: Sphericity / Irregularity**

It is known that irregularly shaped lesions are more likely to be malignant by virtue of consisting of tissues that grow irregularly and often invasively. Benign lesions are more likely to be encapsulated and thus exhibit spherical shapes. We have elected to measure the irregularity of a lesion by sphericity, defined as:
Where, \( V_{\text{lesion}} \) is the volume of our lesion ROI
\( V_{bs} \) is the volume of the bounding sphere that fully encompasses our ROI

\[ S_p = \frac{V_{\text{lesion}}}{V_{bs}} \]  

*Feature Measurement #4: Average Edge Diffuseness*

The motivation behind this measurement is that malignancies tend to grow into neighbouring tissues thus they tend to exhibit diffuse edges on magnetic resonance breast images. Benign lesions tend to be encapsulated and as such tend to exhibit sharp edges on magnetic resonance breast images. This measurement is made by exhaustively selecting two neighbouring voxels, one of which is in our ROI (\( iR \)) and one of which is not (\( oR \)). Thus we compare the signal intensity difference between the border pixels of our potential malignancy and any neighbouring pixels belonging to another group. This difference value is normalized by the dynamic range of the data. This computation for a single pair of neighbouring pixels is formalized by the following equation which is computed on the final time point from the subtracted data set:

\[ \text{PP}_{\text{Diff}} = \frac{|iR_{FTP} - oR_{FTP}|}{dR_{FTP}} \]  

Where, \( FTP \) is the final time point of the examination
\( dR_{FTP} \) is the maximum SI minus the minimum SI of the final time point

The average of the collection of calculated PP\(_{\text{Diff}} \) values (from equation 9) is computed to form the average edge diffuseness feature measurement.
Feature Measurement #5: Average Heterogeneity

It is known that cancerous lesions are often characterized by irregular vasculature and so can exhibit heterogeneous signal intensities after administration of the contrast agent. Benign lesions tend to be characterized by a more uniform vasculature and so may exhibit less heterogeneous signal intensity values. Computation of the heterogeneity measurement is made by exhaustively selecting two neighbouring voxels \((iR_1, iR_2)\), both of which are in our ROI. For each voxel pair the following value is computed which takes the signal intensity (SI) difference for the final time point from the non-subtracted data set:

\[
PP_{\text{Diff-Hetero}} = \frac{|iR_{1FTP} - iR_{2FTP}|}{dR_{FTP}} \tag{10}
\]

Where, \(FTP\) is the final time point of the examination

\(dR_{FTP}\) is the maximum SI minus the minimum SI of the final time point

The average of the collection of calculated \(PP_{\text{Diff-Hetero}}\) values from equation 10 is computed to form the average heterogeneity feature measurement.

3.1.1.K Methods – Segmentation – Statistical Analysis

For each of the measurements addressed above variations in the amount of separation obtained between the malignant and benign lesions was tracked for a range of realistic input parameter settings for the proposed segmentation method and for the enhancement threshold segmentation method. The separation between measurements of malignant and benign lesions was measured through receiver-operating characteristic (ROC) curve analysis [78].
3.1.2 Results – Segmentation

We have applied the aforementioned region-of-interest identification technique to a fat-saturated magnetic resonance image of the breast at a range of Euclidean distance parameter settings (5, 15 and 25%) to illustrate the behavior of the program with changes to this input parameter (see Fig. 3.5). The top left of the figure contains the original breast image (with a cancerous lesion in the upper central area). The upper right image provides the resultant regions-of-interest obtained at a Euclidean distance parameter setting of 5%. Red lines mark the borders between adjacent regions-of-interest, this ROI visualization method was selected for the ease of qualitative evaluation. The lower left and lower right images provide the resultant ROIs at the 15 and 25% Euclidean distance parameter settings respectively. For these results the custom metric’s threshold parameter setting of 0.6 was unchanged (as applied to Eq. 4).
Fig. 3.5. A Fat-saturated MRI image of the breast (upper left), segmented at 5% (upper right), 15% (lower left) and 25% (lower right) Euclidean distance parameter settings. Red lines mark borders between adjacent regions-of-interest.

For each of the measurements addressed in the Methods (section 3.1.1.J), the separation obtained is measured by receiver-operating characteristic curve analysis [78]. The ROC area is computed for both the proposed segmentation method and the enhancement threshold segmentation method as the input parameters for each technique are varied. The resultant ROC area measurements for each of our five measured features (mean slope,
mean washout, irregularity, edge diffuseness and heterogeneity) are provided in Fig. 3.6. The ROC area is different for each feature measurement and also varies based on the setting of the given segmentation approach addressed.

**Fig. 3.6.** The variation in the ROC curve area for each segmentation technique applied to five feature measurements which each produce their own ROC areas. The lower x-axis provides the parameter settings for enhancement threshold segmentation, the upper x-axis provides the parameter settings for the proposed segmentation’s equation 4 parameter. Black arrows indicate the appropriate x-axis scale for each segmentation technique.
The best separation obtained at fixed input parameter settings for each of the two segmentation techniques for each of the five features is provided in Fig. 3.7.

**Fig. 3.7.** Comparative ROC areas for each of the five feature measurements at appropriate segmentation parameter settings (proposed technique’s threshold (eqn. 4): 0.62, enhancement threshold: 95%). Each unique measurement produces a unique ROC area. Confidence interval (95%) spreads are provided as computed by SPSS (SPSS Inc., Chicago, IL, USA).

We have also provided an example segmented magnetic resonance breast image in Fig. 3.8, whereby red lines mark borders between adjacent groups.
Fig. 3.8. A subtraction breast MRI image with an invasive ductal carcinoma exhibiting diffuse edges (above). Blue lines mark the segmentation regions extracted by the proposed method. ROIs with a mean washout greater than 0.12 are colour coded red as potentially cancerous.
3.1.3 Discussion – Segmentation

This thesis chapter presents a new methodology for identifying regions-of-interest in images. We have qualitatively evaluated the technique by running it on a variety of image types with widely varying applications and provided the final regions-of-interest in the results. Many of the most established segmentation techniques (k-means, c-means) take an input parameter that defines the number of resultant groups in the image that the program is tasked to find. This approach is useful in applications where we know how many resultant groups we want in the final image. However, in many applications (such as screening for cancer from MRI, or identifying weather patterns from satellite images, or identifying pathologies from cell histology slides) we inherently don’t know how many ROIs should be in the final segmented image. One of the strengths of this approach is that the resultant number of ROIs is flexible. When the input parameter is set low, the image’s ROIs will be small and thus the overall segmentation will identify many ROIs. With the input parameter set high the image’s ROIs will be larger and thus the overall segmentation will produce considerably fewer ROIs. Because our technique avoids making assumptions regarding the underlying data it is both robust and potentially widely useful in a variety of application fields.

For the purposes of the study in this thesis chapter, two neighbouring groups are merged if the computed metric from equation 4 is greater than a threshold value. This threshold value was varied from 0.55 to 0.65 in steps of 0.01. It was noted that when this threshold value was lowered or raised out of this range, that the segmentation technique did not successfully segment all of the lesions in the dataset. Furthermore the process of
segmenting 223 examinations at many more segmentation parameter settings is extremely time consuming and so was not performed for this thesis (future work will include evaluation of the full range of the technique’s parameter).

It can be qualitatively observed that the proposed technique produces sensible ROIs on a variety of images and image types. In order to validate the efficacy of the approach we need to perform a quantitative analysis. Doing so is challenging as there are no fool-proof methods for quantitatively validating a segmentation algorithm. In the most common quantitative segmentation evaluation method, gold standard regions-of-interest are obtained by manual contouring by expert humans. The resultant segmentations would be compared with those obtained by our proposed method. While this quantitative method is useful for comparing the approach with an expert human, it is known that there is considerable variation between expert humans [87] and as such it may not be wise to consider these expert segmentations a gold standard. Human ROI identification can yield errors due to automatically applying our natural object recognition abilities instead of just basic ROI identification / segmentation. The other quantitative validation method (which is performed in the analyses below) will compare our technique’s (and other existing segmentation methods’) ability to identify ROIs that are useful for allowing the detection of suspicious regions (such as detecting cancer from MRI images of the breast). This quantitative approach sidesteps the issue of needing a pixel-by-pixel gold-standard and instead relies on having an accurate label for an overall region (say as being either cancerous or non-cancerous). This is useful when we do not have gold standard
information regarding the correct location of a border between two regions, as is the case in screening for breast cancer from MRI.

In this thesis chapter, a technique for the segmentation of volumetric multi-spectral images has been presented and the technique has been evaluated on a dataset of magnetic resonance breast images. The standard enhancement threshold based segmentation produced ROC areas consistently lower than those produced by the proposed technique. The proposed technique produces consistently higher or equal ROC areas at an input threshold (equation 4) parameter setting of 0.62 and produces reasonable quality ROC areas across the range of parameter values explored. This helps highlight one of the main advantages of the proposed technique: that it is robust and stable, working well across our full set of feature measures consistently across the range of input parameters explored. This is a highly desirable characteristic in the building of a computer-aided detection program for magnetic resonance imaging of the breast.

The enhancement threshold was selected for comparative evaluation as it is the most widely used technique available [41-42, 48, 62-73]. The second most commonly used technique is known as artificial neural networks (ANNs) [39, 41, 73]. Artificial neural networks have considerable flexibility for the researcher who can change the network’s architecture as well as the learning algorithms performed in each modeled neuron. Because of the widely varying possibilities in the implementation of an ANN and the lack of a recognized gold standard ANN architecture for breast MRI we have elected not to compare our technique with artificial neural networks. Another technique demonstrated
for breast MRI is the mean-shift algorithm [75]. Although the mean-shift algorithm has shown promise it should be noted that our study contains detected cancerous lesions ranging in diameter from as small as 2 mm to as much as 6 cm long. The mean-shift algorithm takes an input parameter known as the spatial bandwidth. Since our lesions vary so widely in size it is unlikely that the mean-shift algorithm will provide sensible segmentations for all of our lesions at the same parameter setting [75] and as such we have elected not to compare our technique with mean-shift. An additional technique applied to the segmentation of breast MRI lesions is fuzzy c-means [77]. It has been shown to be an effective semi-automatic technique (the radiologist must manually identify the sub-region of the examination in order to perform computer-aided diagnosis). However, it is inappropriate for fully automatic ROI detection in this context as appropriate settings for the input parameter (number of groups) are unknown when a screening examination may include a variable number of lesions, non-lesion tissues, partial volume tissues and necrotic tissues.

Although an assortment of alternative feature measures can be found in the literature, typically they are all a measurement of the same type of effect measured by the features we selected (see Methods section 3.2.1.E). Feature Measurements for breast MRI CAD typically measure something related to the permeability of a lesion’s vasculature or the irregularity of a lesion’s shape. We have also included the edge diffuseness measurement (#4). Although we are unaware of a previously presented edge diffuseness feature measure for magnetic resonance imaging of the breast, our technique is somewhat similar to the blooming sign measurement [74]. In our study we were unable to implement the
blooming sign measurement as the temporal resolution of our magnetic resonance acquisition protocol is too coarse. Due to this shortcoming we have elected to implement the outlined edge diffuseness measurement which is performed on the final examination time point. The final time point was selected as benign lesions often exhibit very little enhancement in the early phases of the examination.

All of the measurements taken in this thesis chapter were performed on magnetic resonance breast images obtained using the acquisition protocol identified in the Methods (section II.A). In practice, there are many different MRI acquisition protocols used at imaging centres all over the world. Typical imaging protocol variations include differing spatial and temporal resolutions. The 4x4x2 block size was selected to loosely reflect the x and y spatial resolutions and the in-plane slice thickness of our acquisition protocol. It is expected that if the spatial resolution is significantly changed then the initial 4x4x2 block size may need to be modified as well. This initial block size was selected such that the proposed segmentation technique would be able to segment even the smallest malignant lesions in our dataset. Unfortunately, at present a second independent dataset with significantly differing spatial resolutions is not available and so investigation of this issue is beyond the scope of this study. However, it should be noted that the above formulation was able to automatically segment very small (2 mm) lesions.

It should be noted that the dataset used in this study is both large (223 examinations) and challenging. The data set consists of many ductal carcinoma in situ (DCIS) lesions (typically deemed difficult to detect) and many small lesions (catching lesions as small as
2-3 mm). The main method for detecting cancerous tissues is by use of enhancement and washout features (measurements 1 & 2 in this study). Unfortunately lesions that are small or DCIS often exhibit enhancement and washout values similar to stereotypically benign lesions (ie. not a lot of enhancement and little to no washout). These small lesions also tend to present as simple focal enhancements (somewhat circular/spherical) making their detection by an irregularity/sphericality measure difficult. The proposed edge diffuseness measurements have the potential to assist in detecting some of these challenging lesions. This will be addressed in chapter 4 by combining our set of feature measurements with a multidimensional classifier such as support vector machines which have been previously demonstrated to be a promising technique for the classification of breast MRI lesions [51]. This final classification step will combine our set of features and make final predictions as to whether a given lesion is malignant or benign thus forming a complete CAD system for MRI of the breast. This classification / supervised learning extension will require thorough evaluation through validation as well as appropriate feature selection, and is addressed in chapter 4.
Chapter 4

Classification: Dynamic and Shape Information
This thesis chapter is concerned with supervised learning (or classification). Supervised learning is a pattern recognition technique that provides a set of formal mathematical rules that assign a given state (say cancer) based on a set of labeled training data. These rules are formalized into a set of computational instructions. In this chapter, the final classification step of a computer-aided diagnosis and detection system is addressed while combining both dynamic and shape based feature measurements. The purpose of this thesis chapter is also to present a new bias-similarity based supervised learning technique, demonstrate how it facilitates receiver operating characteristic curve analysis and to comparatively evaluate its performance against the established support vector machine method.

4.1 Bias-similarity based supervised learning

The purpose of this thesis section is to present a bias-similarity based supervised learning algorithm and to demonstrate how it facilitates receiver-operating characteristic (ROC) curve analysis. ROC curve analysis evaluates the tradeoff between the sensitivity and specificity of a test and is used extensively not only in medical imaging but in testing theory in general. ROC curve analysis is facilitated by creating a supervised learning technique with a single parameter that biases the test towards one of our two groups, thus ROC analysis can be performed by simply varying the proposed method’s only input parameter. Supervised learning is performed by combining our bias input parameter with a similarity measurement.
4.1.1 Methods – Classification: Dynamic and Shape Information

4.1.1.A Methods – Supervised Learning – Image Acquisition
The screening protocol used is identical to that used in the thesis section 3.1.1 (44 malignant and 179 benign).

4.1.1.B Methods – Supervised Learning – Image Registration
Image registration is identical to that used in the thesis section 3.1.1 [81].

4.1.1.C Methods – Supervised Learning – Lesion Measurements
For each of the 223 lesions addressed in this study 4 features were measured (lesion time normalized enhancement, lesion enhancement reduction, lesion irregularity and lesion diffuseness – see section 3.1.1 for their definitions). This yields a four dimensional data space within which malignancies will be predicted by a multidimensional classification (or supervised learning) technique. Each of these features has been scaled to the range 0 to 1.

4.1.1.D Methods – Supervised Learning – Proposed Classification Technique
A classification technique operates by inputting a set of training data with known labels/classes. A classification technique will then be able to predict an unknown test sample as belonging to one of the provided labeled groups by evaluating it with respect to the provided training data. The research and development of our proposed supervised learning / classification technique involves the main motivating idea that the technique should take only a single input parameter that biases the test towards one of the two
groups. It was then determined that high quality results were obtained when we combine the sample biasing feature with a similarity measurement. For our similarity measurement we selected the sum of the distances between our test sample and all of our training samples, which is an inverse similarity measure (large numbers indicate dissimilarity). The proposed technique will predict the class of a given sample as defined in equation 11 as follows:

\[
\text{Class} = \text{sign}\left(\alpha \frac{\text{positiveCount}}{\sum_{i=1}^{\text{samples}} \text{label}(i) \sqrt{\sum_{j=1}^{\text{dims}} (\text{trainSet}(i, j) - \text{test}(j))^2}}\right) - (1 - \alpha) \frac{\text{negativeCount}}{\sum_{i=1}^{\text{samples}} (1 - \text{label}(i)) \sqrt{\sum_{j=1}^{\text{dims}} (\text{trainSet}(i, j) - \text{test}(j))^2}} \right)
\]

Where, \(\text{samples}\) is the number of training data samples
\(\text{dims}\) is the number of measurements per sample
\(\text{trainSet}\) is the input training data
\(\text{test}\) is the input vector to be tested
\(\text{sign}(x)=1, x \geq 0; \text{sign}(x)=-1, x<0\)
\(\text{positiveCount}\) is the count of positive samples
\(\text{negativeCount}\) is the count of negative samples
\(\text{label}(i) = \begin{cases} 1 & \text{if sample } i \text{'s label is positive} \\ 0 & \text{if sample } i \text{'s label is negative} \end{cases}\)
\(\alpha\) is the input bias parameter with a range from 0 to 1

This equation’s behaviour is illustrated in Fig. 4.1 which demonstrates the effect of varying the equation’s only input parameter: \(\alpha\). The equation is such that tuning the \(\alpha\) parameter high will bias the test in favour of the positively labeled samples because the \(\alpha\) term is multiplied by the inverse mean distance to the positively labeled samples.
Similarly the inverse mean distance to the negatively labeled samples is multiplied by \(1-\alpha\) which biases the test towards the negative samples when \(\alpha\) is set low. Fig. 4.1 demonstrates how the resultant decision border is affected by the input parameter’s setting.

![Proposed Supervised Learning - alpha variations](image)

**Fig. 4.1.** The behaviour of the proposed classifier with respect to variations in the input \(\alpha\) parameter (0.45; 0.5; 0.55).

The proposed formulation is very similar in operation to support vector machines without an optimization step that down-samples and assigns weights to the training data (support vectors). Instead our technique can be thought of as a support vector machine with every sample included as a support vector in an effort to maintain statistical precision. The weight on each support vector is simply the Bias term in equation 11 (\(\alpha\) and \(-(1-\alpha)\) for the two classes respectively), thus making the weights different for the positive and
negative classes. The similarity measure (the inverse of the mean of the sum of the
distances in this case) is analogous to a support vector machine kernel function.

The above formulation adjusts for situations when we have a differing number of samples
in each of our two groups by normalizing the inverse of the sum of the distances (the
similarity measure) by the number of positive and negative samples respectively. Of
course the input bias term ($\alpha$) can account for this effect on its own, however, the
appropriate setting of the bias term will be related to the relative number of samples in
either group. By scaling the similarity measure by the number of samples in each group
we can ensure that sensible settings for the $\alpha$ parameter are found around the setting 0.5.

This formulation (equation 11) involves a classification function that divides two
classes/groups. For instances when a problem has more than two classes the equation can
still be used but will have to be repeated for each class to be tested with the positive label
assigned to the class currently being tested and a negative label assigned to all other
samples. The breast magnetic resonance imaging computer-aided diagnosis examples
used below consist of only two classes (malignant and benign lesions).

It was found that when using the proposed supervised learning technique (or most
supervised learning techniques) each feature measurement should be scaled to a fixed
range (in this paper scaling was performed from 0 to 1). This helps ensure that a single
measurement does not dominate the decision process as compared with other
measurements with inherently smaller scales.
In order to comparatively evaluate our proposed segmentation technique, we elected to compare our technique with support vector machines which have been shown to be the top performing classification technique for breast MRI when compared with SER, density based classifiers (like k-nn), and others [51, 52]. Additionally, we have elected to compare our proposed classification technique with support vector machines because of its similarity to SVMs which is addressed in more detail in the discussion.

Support vector machines are an established machine learning and pattern recognition technique [82] widely used in research and were introduced in chapters 1 and 2. For the purpose of this study, we are looking at a radial basis function (RBF) kernel (see equation 3 in section 2.1.1) which has been previously shown to be the best performing kernel function in many applications including CADx for breast MRI [51, 52]. Non-linear kernel functions such as the radial basis function (equation 3) provide flexibility to the SVM classification function calculation. These equations allow the definition of the classification function to be non-linear in the input space. If the kernel transformation function does not fully separate our data, a slack error variable is used to create a soft margin classification function for data separation. Support vector machine based classification has been implemented using the *libsvm* open source library [83].
4.1.1.F Methods – Supervised Learning – Validation and Statistical Analysis

For both the proposed technique and for the RBF SVMs we perform leave-one-out cross-validation. The radial basis function support vector machine has its input parameter varied according to the following equation:

\[ \gamma = e^{\Gamma} \]  \hspace{1cm} (12)

where \( \Gamma \in \{-7.0, 4.5\} \) in steps of 0.1. For each setting of the input parameter \( \gamma \), the area under the receiver operating characteristic curve [78] (ROC) is computed. For the SVM technique this is performed by the standard method; the classifier is run in regression mode and the threshold applied to these regression values is varied to produce points along a ROC curve. For the proposed technique a single receiver operating characteristic curve is computed by varying the input parameter \( \alpha \) from 0 to 1 in uniform steps of 0.005.

The second validation technique addressed is randomized trials. Whereby 50% of the samples in our data set were randomly selected and assigned for use in training. The remaining 50% of the samples were used for testing. This random selection process was repeated 100 times and both of the proposed classification techniques and support vector machines were compared.

4.1.2 Results – Classification: Dynamic and Shape Information

For the combined set of measurements addressed in the Methods (sub-section C), the separation obtained between malignant and benign lesions is measured by receiver-operating characteristic curve analysis. Changes in the ROC area for leave-one-out
validation are tracked with variations in the input parameters for the support vector machine radial basis function technique. A single ROC area was computed for the proposed technique (equation 11). These ROC area calculations were repeated for the randomized trial experiments and the plot of how the standard deviation of the ROC areas from the randomized trials is provided in Fig. 4.2.
Fig. 4.2. Comparative results of the proposed method and RBF SVMs for leave-one-out (top) and randomized trials (middle) based validation. The standard deviation of the ROC areas from the randomized trials are also provided (bottom).

We have also provided classifier visualization plots [51] to assist in qualitative understanding of the effects of the input parameters on the classifying decision function.
and to assist with a general understanding of the two techniques being compared (see Fig. 4.3 and 4.4).

**Fig. 4.3.** Principal component space projection plots with classifier boundaries for the proposed method (above, alpha values provided) and the SVM method (below, $\gamma = 0.12$, three bias threshold boundaries provided (left to right): -0.89, -0.87, -0.82).
Fig. 4.4. Principal component space projection plots with classifier boundaries for the SVM method demonstrating over-fitted solutions (upper pane, $\gamma = 49.4$) and the SVM method demonstrating linear-like behaviour for low $\gamma$ values (lower pane, $\gamma = 0.001$). Red dots mark malignant measurement, green dots mark benign measurements.

4.1.3 Discussion – Classification: Dynamic and Shape Information

We have presented a new bias-similarity technique for performing supervised learning and compared it with the established support vector machines on a set of breast MRI data. It should be noted that the proposed technique is entirely data-driven and as such should also perform well in applications outside of breast MRI CADx. One of the strengths of the proposed technique is its robustness and ease-of-use. Furthermore, the validation
results presented in this thesis chapter demonstrate that the easy to use single-parameter proposed technique performs equivalently to the established support vector machine method, which requires two parameters to be set. The support vector machine technique required a full exploration of the unbounded $\gamma$ parameter (as previously mentioned $\gamma = e^{\Gamma_{\text{Max}}}$, where $\Gamma_{\text{Max}} \in \{-7.0, 4.5\}$). For each of these $\gamma$ settings a ROC area calculation needs to be performed (in the form of varying a threshold applied to the regression results of the support vector machine classifier for any given $\gamma$ value).

Another advantage of the proposed technique is how it simplifies receiver operating characteristic curve area calculations. ROC areas can be easily computed by simply varying the $\alpha$ parameter across its bounded range of 0 to 1 and recording the sensitivity and specificity values for any given $\alpha$ setting. ROC area calculations are extremely common in computer-aided diagnosis applications and in testing theory in general.

The radial basis function based support vector machine’s $\gamma$ parameter is described as the radius of each support vector which in turn affects the curvature of the separating classification function. This causes the radial basis function support vector machine to exhibit linear classifier behaviour (underperforming) as $\gamma$ approaches 0 (see Fig. 4.4 B) and produces consistently over-fitted solutions as $\gamma$ gets very large (see Fig. 4.4 A). Good solutions are typically found at parameter settings somewhere in between these values and this unbounded parameter space needs to be thoroughly explored to ensure the generation of high quality results. While support vector machines have proven themselves in many applications with the radial basis function [34, 36, 51-52], having a parameter that affects the curvature of the resultant decision function is counterintuitive.
since the researcher will not typically know appropriate curvature settings of their classification function a priori. In this thesis section we are arguing that the curvature of the decision function should not be controlled directly by a user parameter, but instead should be implicitly derived from the training data distribution. It should also be noted that the standard deviation of the receiver operating characteristic curve area is quite variable for the support vector machine radial basis function technique.

A common use of classification functions such as support vector machines is to use them in regression mode. When predicting in regression mode a classifier will output a number instead of the typical +1 or -1 to indicate which of the two groups it belongs to. Regression prediction can be used as a measure of how far the given sample is from the decision function. The proposed technique (equation 11) can support regression by simply removing the \textit{sign} function from the equations.

Although the proposed technique consists of just a single parameter that controls the bias towards either of our class data, the technique is not limited to such simple formulations. Datasets may arise where a different similarity measure that takes in an additional input parameter may be desirable in order to produce the highest quality solutions possible. A simple example would be to include a second parameter, for example $\beta$, which controls the percentage of nearest training samples from either class that are used in the prediction of a new test sample. This would result in the prediction process behaving more like support vector machines in that the prediction is based predominantly on those samples nearest to the test sample. This technique would behave as a nearest neighbour based
classifier when the parameter $\beta$ approaches 0 and (and $\alpha = 0.5$), and would behave the same of the main proposed technique as $\beta$ approaches 100% (all $\alpha$ settings). The proposed technique weights each vector equally based on the bias parameter. A k-nearest neighbour classifier heavily weights samples close to our test sample and weights the rest of the samples 0 which can easily lead to overfitting. This $\beta$ parameter would be relevant in that it directly controls the amount of local training data used in the decision process. It was determined that this dual parameter formulation of bias-similarity was inappropriate for use in this study due to the limited number of samples available.

The main disadvantage of the proposed technique is that the decision process is slower than that of support vector machines by virtue of having every data sample contribute to the decision function. The degree to which the proposed technique is slower is largely dependent on the number of support vectors selected in the SVM learning process. In the data sets used in this thesis it would be fairly common for SVMs to downsample to about one third of the total data samples, thus the proposed technique may perform 3 times as slowly in such a situation. This is equivalent to having every data sample being a support vector. In support vector machines, prediction is accomplished by combining the test vector with each of the support vectors via the kernel function. The same is true with our proposed technique, but all of the training samples are included.

The proposed technique yields ROC areas around 0.80 to 0.81 depending on the validation technique selected. This can yield a prediction test by the proposed method that achieves a sensitivity of 89% and a specificity of 70%. Unfortunately comparing
these results with those obtained by radiological analysis is not possible because many of
the lesions in our dataset were rated BI-RADS 0 (needs further workup) by the
radiologist. Furthermore, enhancing tissues not identified as suspicious by the radiologist
were ignored. Evaluation of the rate of false positive predictions per examination for any
of the classifiers discussed was not possible due to the nature of a magnetic resonance
breast examination. In cases where a benign diagnosis has been determined by one year
disease free follow-up we can assume that all lesions in the breast are negative, however,
when the lesion status has been established by biopsy, we cannot know the status of any
additional lesions that have not been sampled.

In the situation where a set of new training samples are added to an existing training set, a
number of technical issues arise in the use of supervised learning algorithms. In support
vector machines the addition of new data would necessitate retraining the SVM. With the
proposed method there is no need to retrain the system as no training occurs. If the added
samples fall on the outskirts of the two groups then effectively no changes will occur to
the separating classification function by the proposed technique. However, if the samples
fall near the existing separating classification function, then the equation of this
separating decision function will shift somewhat. It is possible that the most appropriate
value for the input parameter ($\alpha$) will need to be re-evaluated and previous validation
results will not be applicable.

It was already known from Chapter 2 that enhancement features were useful for lesion
discrimination. If we combine just the 2 enhancement features from section 3.1.1 we get
reasonable ROC areas. If we add the irregularity feature and the edge diffuseness feature to the enhancement features we get further ROC area improvements, however, observed improvements tend to be small (amount of improvement is dependent on which validation technique is being addressed). Combining the heterogeneity feature with the remaining features does not improve the ROC area results, suggesting that whatever separation this feature provides is already separated by the previous set of measurements.

The validation methods selected for classifier evaluation in Chapters 2 and 4 are leave-one-out and randomized trials validation. These were selected as they are standard evaluation methods. The leave-one-out technique is a form of jackknifing, whereby the data set is divided into a number of equally sized groups and then leave-one-out validation is performed on the groups. In this thesis I elected to perform the jackknifing technique whereby each individual sample is assigned its own individual group for the leave-one-out validation trials. This jackknifing technique was selected as our data set was small enough to perform this exhaustive validation within a computationally feasible time frame. The randomized trials validation method involved randomly selecting 50% of our data set for training and reserving the rest of the data for testing. This random process was repeated many times. This is very similar to the bootstrapping technique except that the randomized technique selected does not allow multiple data samples to recur in a single training set. It should be noted that our data set is small and so performing validation based on the similar bootstrap may be beneficial.
The main novelty of the classification approach presented in this thesis chapter is that it is a simplified reformulation of the idea behind support vector machines. Support vector machines are a fairly complicated technology which implement both a training and prediction process whose behaviour can be modified based on a number of parameters. The idea behind support vector machines is that they construct an implicit decision function out of a weighted subset of the training samples. The proposed technique is both simple and novel in that it mimics the overall effect created by support vector machines but does so in a simple way that incorporates all of the training data available (which can help ensure statistical reliability). This also eliminates the need for a training phase since all the data is used, however a down-sampling phase may be added in the future in order to improve the technique’s execution speed. Furthermore, the proposed technique only contains a single user parameter (bias – $\alpha$ that is bounded from 0 to 1) whereas a high performing support vector machine based on the radial basis function contains two parameters ($\gamma$ which is unbounded and affects the curvature of the resultant decision function and a less elegant bias term).
Chapter 5

Conclusions and Future Work
5.1 Conclusions

This thesis has presented new computational and pattern recognition techniques that can be used in many fields. A new supervised learning (classification) technique has been presented, as well as a method for visualizing the resultant classification decision function. A new unsupervised learning (segmentation/clustering) technique has also been presented. All of the new techniques developed contribute to an overall computer-aided detection and diagnosis system for breast cancer from MRI examinations. The developed pattern recognition techniques have been evaluated by having their ability to assist in discriminating between malignant and benign lesions assessed.

A final computer-aided diagnosis and detection (CAD) system is represented by the block diagram provided in Fig. 1.7 (page 18) and is comprised of a combination of the pattern recognition techniques presented in this thesis. The final CAD system inputs a breast MRI examination. The examination is then divided into regions-of-interest by the proposed volumetric multi-spectral segmentation technique (see chapter 3). For each of the resultant regions-of-interest a set of dynamic and shape measurements are performed. Ideally, all of the regions-of-interest from a testing examination are compared for possible malignancy by the bias-similarity based supervised learning technique.

The proposed image segmentation technique (chapter 3) has a range of potential applications. In order to demonstrate this we have provided Fig. 5.1 which contains 4 images each segmented at the exact same segmentation parameter setting. The upper left
image is a computed tomographic (CT) image of the chest obtained from the MedPix Medical Image Database (http://rad.usuhs.edu/medpix/medpix_home.html). The upper right pane is a satellite image of cyclones in the Pacific ocean obtained from the NASA visible earth project (http://veimages.gsfc.nasa.gov/). The lower left pane is a colon cell histology image from the online information centre for immunohistochemistry (http://ihcworld.com/imagegallery/). The lower right pane is an MR image of the brain.
Fig. 5.1. A montage of a CT image of the chest (upper left), a satellite image of cyclones in the Pacific (upper right), a histology image of colon cells (lower left) and an MRI image of a brain (lower right) each segmented by the proposed method at the same threshold setting for equation 4 (0.6). Red lines mark borders between adjacent ROIs.
The maximum ROC area obtained by combining the features presented was 0.84, providing a test potentially yielding a sensitivity of 89% and a specificity of 70%. This research has the potential to improve breast imaging by providing an accurate automated second reader to assist radiologists in the diagnosis of breast cancer from MRI, thereby reducing the variability common in the analysis of the large amounts of data in a breast MRI examination [32]. Comparing these results with those of a radiologist is challenging as the radiologist had the ability to order a follow-up imaging examination if they felt that the screening images were not clear enough. This diagnosis is referred to as BI-RADS 0 and was the diagnosis in over half the examinations used in this thesis. The CAD system does not have the ability to order a follow-up exam and future work will look at ways to combine follow-up examinations into the overall CAD process. If one were to evaluate the performance of the radiologist based exclusively on the screening data (while marking a BI-RADS 0 classification as a positive diagnosis – as is common in the literature) then the radiologist yields a sensitivity of 89% and a specificity of 63%. However, if one were to evaluate the performance of the radiologist based on the combination of the screening and follow-up data then the radiologist has a ROC area of 0.82 yielding a sensitivity of 84% and specificity of 86%. While the radiologist outperforms the CAD technique in terms of specificity, the CAD technique outperforms slightly in terms of sensitivity. Furthermore, the CAD technique produced these results without having the advantage of taking measurements from the follow-up examinations. Another general assessment can be performed by comparing these CAD results with the results obtained by a set of 15 radiologists tasked with all analyzing a set of breast MRI examinations [32]. It should be noted that the CAD results were exclusively generated
with breast MRI data obtained at Sunnybrook Health Sciences Centre, whereas the 15 radiologists were analyzing examinations performed in the United Kingdom. The screening process from the set of 15 radiologists was determined to have a ROC area of 0.85, sensitivity range of 77-95% and a specificity range from 61-77%. Similar results were obtained by our CAD system (ROC area 0.84, sensitivity 89%, specificity 70%).

In spite of the inability to directly compare the CAD results with those obtained by a trained radiologist, a number of other aspects of this CAD research should be highlighted. The first CADx system built was presented in chapter 2 and was shown to outperform the methods used by existing commercial breast MRI CADx systems (Confirma Inc., Kirkland, WA, USA - Sentinelle Medical Inc., Toronto, ON, Canada) and has been published in *IEEE Transactions on Medical Imaging* [51] and in *Academic Radiology* [79]. In chapter 3, a segmentation technique was presented that allows for the extraction of shape measures that can further improve the CADx system presented in chapter 2. In chapter 4 these extra measurements are combined with a new simpler formulation of the methods used in chapter 2. This yields a final CAD system with a ROC area of 0.84 which outperforms any of the CAD techniques presented in chapter 2 (which outperformed existing commercial techniques). Comparative evaluation and validation of the CADx system presented in chapter 4 is the subject of future work.

The results of this CAD system look promising when compared with a radiologist who has the ability to refer to follow-up examinations. The CAD results also look promising when compared to those obtained by a set of 15 radiologists (looking at a different data
set). Finally the CAD results presented in this thesis outperform the CAD results obtained by methods in use in existing commercial products. It is known however, that we can get some CAD improvements by including more information (high resolution follow-up examinations or including additional MRI measurement techniques like the apparent diffusion coefficient etc.). Ideally a final CAD system will incorporate any available information that can improve the overall diagnostic process.

5.2 Future Work

Comparative evaluation from chapters 3 and 4 was only performed on the radiologically identified lesions, thus ignoring enhancing tissues not identified as suspicious by the radiologist. Evaluation of the rate of false positive predictions per examination for any of the classifiers discussed was not possible due to the nature of a magnetic resonance breast examination. In cases where a benign diagnosis has been determined by one year disease free follow-up we can assume that all lesions in the breast are negative, however, when the lesion status has been established by biopsy, we cannot know the status of any additional lesions that have not been sampled. In order to form firm conclusions regarding the robustness of the test, performing a validation study on an independent data set will be necessary. The work presented in this thesis is based on a single breast MRI surveillance database collected over more than 10 years at Sunnybrook Health Sciences Centre (Toronto, ON, Canada). A study comparing multiple databases will allow us to investigate unanswered questions such as how transferable are the outlined techniques to different MRI acquisition protocols. It is desirable to acquire as many breast MRI
examinations as possible to form firm conclusions about the results presented, fortunately a new multi-centre database is being planned. The separation between malignant and benign lesions obtained (quoted above, see section 5.1) will unfortunately result in many benign lesion biopsies (179 benign lesions with 70% specificity or a 30% biopsy rate yields 54 benign biopsies). Ideally we would be able to build a CAD system that catches all of our cancers while not causing the biopsy of too many benign lesions.

Based on both quantitative and qualitative measures improving the spatial resolution and the signal-to-noise ratio of breast MRI imaging would probably be beneficial. Such improvements will allow us to make more accurate shape measurements and to avoid noise in the measuring process. A study whereby every single follow-up hi-resolution examination is analyzed for shape features and correlated by position (ideally automatically) to the enhancement data would be beneficial. It should be possible to compare the shape measurements obtained from the hi-resolution fat saturated follow-up images to the shape measurements obtained from the same lesion imaged by the screening protocol. This type of analysis could help determine if shape measures can provide us with more useful discriminating features at higher resolutions. If the answer is yes (shape measures improve significantly with higher resolution) then firm recommendations regarding researching higher resolution MRI protocols can be made. Recently there has been quite a push in the MRI physicist’s community towards breast imaging protocols that acquire more volumetric images. Improvements to the temporal resolution are thought to be beneficial in assisting in accurate measurements of the signal intensity time curve (and can also assist in pharmacokinetic modeling). However, the
curves themselves are fairly simple. It is questionable whether temporal resolution improvements will actually assist in further delineation between difficult to classify malignant and benign lesions. At present the screening program at Sunnybrook hospital has caught multiple small (diameter: 2-3mm) tumours. Since a small tumour such as this may consist of as little as 10 pixels/voxels, it seems plausible that further improvements will be found by increasing the spatial resolution of the images.

Although we have comparatively evaluated our region-of-interest identification technique against the established enhancement threshold based segmentation, performing a study that will look at validating our segmentation results against those created by an expert radiologist would be beneficial. Human error factors into such an experiment, but the results can be useful to determine how much the algorithm deviates from a set of expert human observers. Comparison with expert humans was outside of the scope of this study as it is not feasible to expect our radiologists to manually segment 223 lesions due to time constraints. Future work will look exclusively at this issue.

The segmentation process’ D metric (equation 4) will also be subjected to traditional statistical analyses such as demonstrating the distribution of the D value of a large set of randomly generated distributions and by performing a Monte Carlo simulation.

It should be noted that the feature measurements addressed in Chapter 3 involve averaging together the computed values for each pixel/voxel within a given region-of-interest. It is known that the most enhancing sub-region of an ROI can be a useful marker
for detecting malignancies, as they are often characterized by inhomogeneous vasculature. Although the study presented in chapter 1 was performed on a pixel-by-pixel basis, the features from chapter 3 were spatially averaged together. In practice, isolating the most enhancing pixel/voxel did not provide separation between our malignant and benign lesions as reliably as the averaged measures. It is believed this was due to noise levels on the subtraction images from which these measurements were taken. Future work will look at the inclusion of more features, including texture features of varying scales and derivatives of the signal-intensity time series. Once many features have been computed, feature selection can become a significant issue. On this topic we have been collaborating with Dr. Don McGaughey at Royal Military College of Canada (Kingston, ON) on the use of fast orthogonal search (FOS), a method for assisting in the selection of relevant measurements from a large set of features. This thesis was focused more on classifier research, evaluation and development than feature selection. Although it is expected that small improvements can be found through superior classification techniques, larger CAD improvements can probably be found through feature selection.

All of the evaluations presented in this thesis were devoted to assessing the performance of the CAD system. However, in practice such a system would be used by a radiologist who is responsible for making the final diagnosis of the enhancing lesions seen on a given breast MRI examination. How the interaction between the CAD system and the radiologist affects the resulting radiological diagnoses is unknown. Performing a study that would address this problem by having two trained radiologists interpret a set of normal breast MRI images, and CAD processed breast MRI images would be beneficial.
A full data set can be CAD processed and split into two halves. The first radiologist receives the full data set but with half the breast MRI images unprocessed and the other half CAD processed. The second radiologist receives the same dataset but with an opposite selection of unprocessed and CAD processed images. Thus each of our radiologists will diagnose a set of unprocessed images as well as a set of CAD processed images evaluated by the other radiologist in an unprocessed manner. Such a study could provide us with valuable information as to the effect of diagnosing breast MRI examinations with the help of a computer-aided detection and diagnosis system.

Finally, in the context of a patient’s cancer screening, typically considerably more information is available about the patient than merely the breast MRI examinations themselves. Sometimes the patient is also imaged by mammography, where we might have the radiologist’s diagnosis of the mammogram and we may even have mammographic CAD processed results available as well. An ultrasound breast examination may also have been performed. Furthermore, we may have information about the patient’s medical history which may also affect the likelihood of the MR examination containing a cancerous lesion. Future work will look at methods for combining available patient data and all available cancer screening data (MRI, mammography, ultrasound, etc.) into a complete CAD system.
References


[38] A. Degenhard, et al., “Comparison between radiological and artificial neural network diagnosis in clinical screening,” 


