THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

School of Information Technology
& Electrical Engineering

# Computer Aided Analysis of Dynamic Contrast Enhanced MRI of Breast Cancer

---

**Yaniv Gal**

*A thesis submitted for the degree of Doctor of Philosophy at*

*The University of Queensland in May 2010*

*School of Information Technology and Electrical Engineering*
*The University of Queensland*
*Australia*

*To my son, Orr*

# Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the Copyright Act 1968.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material.

Yaniv Gal: _____

Date: _____

# Statement of Contributions to Jointly Authored Works Contained in the Thesis

## Refereed journal papers

1. Y. Gal, A. Mehnert, A. Bradley, K. McMahon, D. Kennedy and S. Crozier, “Denoising Dynamic Contrast Enhanced MR Images Using Dynamic Non-Local Means”, IEEE Transactions on Medical Imaging, vol 29(1), pp.302-310, 2010.

| Category | Y. Gal | A. Mehnert | A. Bradley | D. Kennedy | K. McMahon | S. Crozier |
|---|---|---|---|---|---|---|
| Analysis and interpretation of data | 60% | 20% | 10% | 5% | 5% | |
| Conception and design | 60% | 30% | 20% | | | |
| Drafting and writing | 50% | 30% | 10% | | | 10% |

2. Y. Gal, A. Mehnert, A. Bradley, D. Kennedy and S. Crozier, “Automatic classification of suspicious lesions in DCE-MRI of the breast”, submitted to Elsevier Artificial Intelligence in Medicine.

| Category | Y. Gal | A. Mehnert | A. Bradley | D. Kennedy | S. Crozier |
|---|---|---|---|---|---|
| Analysis and interpretation of data | 60% | 20% | 10% | 10% | |
| Conception and design | 50% | 30% | 20% | | |
| Drafting and writing | 50% | 30% | 10% | | 10% |

## Refereed conference papers

3. Y. Gal, A. Mehnert, A. Bradley, K. McMahon, and S. Crozier, "An evaluation of four parametric models of contrast enhancement for dynamic magnetic resonance imaging of the breast," IEEE Engineering in Medicine and Biology Society (EMBC), Lyon, France, 2007.

| Category | Y. Gal | A. Mehnert | A. Bradley | K. McMahon | S. Crozier |
|---|---|---|---|---|---|
| Analysis and interpretation of data | 65% | 20% | 10% | 5% | |
| Conception and design | 60% | 30% | 20% | | |
| Drafting and writing | 50% | 30% | 10% | | 10% |

4. Y. Gal, A. Mehnert, A. Bradley, K. McMahon, and S. Crozier, "Automatic Segmentation of Enhancing Breast Tissue in Dynamic Contrast-Enhanced MR

Images," Proceedings Digital Image Computing: Techniques and Applications (DICTA), Adelaide, Australia, 2007.

| Category | Y. Gal | A. Mehnert | A. Bradley | K. McMahon | S. Crozier |
|---|---|---|---|---|---|
| Analysis and interpretation of data | 65% | 20% | 10% | 5% | |
| Conception and design | 60% | 30% | 20% | | |
| Drafting and writing | 50% | 30% | 10% | | 10% |

5. Y. Gal, A. Mehnert, A. Bradley, K. McMahon, D. Kennedy, and S. Crozier, "A new denoising method for dynamic contrast-enhanced MRI", presented at Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, Vancouver, British Columbia, Canada, 2008.

| Category | Y. Gal | A. Mehnert | A. Bradley | D. Kennedy | K. McMahon | S. Crozier |
|---|---|---|---|---|---|---|
| Analysis and interpretation of data | 60% | 20% | 10% | 5% | 5% | |
| Conception and design | 60% | 30% | 20% | | | |
| Drafting and writing | 50% | 30% | 10% | | | 10% |

6. Y. Gal, A. Mehnert, A. Bradley, D. Kennedy and S. Crozier, “Feature and Classifier Selection for Automatic Classification of Lesions in Dynamic Contrast-Enhanced MRI of the breast”, Proceedings Digital Image Computing: Techniques and Applications (DICTA), Melbourne, Australia, 2009.

| Category | Y. Gal | A. Mehnert | A. Bradley | D. Kennedy | S. Crozier |
|---|---|---|---|---|---|
| Analysis and interpretation of data | 60% | 20% | 10% | 10% | |
| Conception and design | 60% | 30% | 10% | | |
| Drafting and writing | 50% | 30% | 10% | | 10% |

## Symposiums

7. Y. Gal, A. Mehnert, A. Bradley, K. McMahon, D. Kennedy and S. Crozier, “A Variation on Non-Local Means for the Denoising of Dynamic Contrast-Enhanced MR Images“, Symposium on GPGPU Techniques for Medical Image Processing and Simulation, Brisbane, Australia, 2009.

| Category | Y. Gal | A. Mehnert | A. Bradley | D. Kennedy | K. McMahon | S. Crozier |
|---|---|---|---|---|---|---|
| Analysis and interpretation of data | 60% | 20% | 10% | 5% | 5% | |
| Conception and design | 60% | 30% | 20% | | | |
| Drafting and writing | 50% | 30% | 10% | | | 10% |

# Statement of Contributions by Others to the Thesis as a Whole

The advisory team of the thesis (Dr. Andrew Mehnert, A/Prof Andrew Bradley & Prof Stuart Crozier) has performed critical proof reading of the thesis and provided editorial advice as required by the Australian Standards for Editing Practice (ASEP).

# Statement of Parts of the Thesis Submitted to Qualify for the Award of another Degree

None.

# Published Works by the Author Incorporated into the Thesis

None.

# Additional Published Works by the Author Relevant to the Thesis but not Forming Part of it

None.

# Acknowledgements

It would not be possible to get to this point without the help of many people.

First and foremost, I wish to thank my principal advisor, Dr Andrew Mehnert, for his professional guidance and for his support, both professionally and personally, along the way. Also, I wish to thank Andrew for the top-up scholarships and for his help in making my family's landing in Australia easier.

I would like to thank my associate advisors, A/Prof Andrew Bradley, for his guidance and for initially offering me this research project, and Prof Stuart Crozier, for his guidance and for offering me various professional opportunities.

Special thanks to Inbal Gal, Kimberly Nunes and Dr Michael Poole for reading parts of this thesis and providing useful comments.

I acknowledge the Australian Government for providing me with financial support through the Australian Postgraduate Awards (APA) scholarship.

I wish to thank my parents for their encouragement and support throughout my studies.

Last, but by no means least, I thank my wife, Inbal, for standing by my side during this long journey.

# Abstract

This thesis presents a novel set of image analysis tools developed for the purpose of assisting radiologists with the task of detecting and characterizing breast lesions in image data acquired using magnetic resonance imaging (MRI). MRI is increasingly being used in the clinical setting as an adjunct to x-ray mammography (which is, itself, the basis of breast cancer screening programs worldwide) and ultrasound. Of these imaging modalities, MRI has the highest sensitivity to invasive cancer and to multifocal disease. MRI is the most reliable method for assessing tumour size and extent compared to the gold standard histopathology. It also shows great promise for the improved screening of younger women (with denser, more radio opaque breasts) and, potentially, for women at high risk.

Breast MRI presently has two major shortcomings. First, although its sensitivity is high its specificity is relatively poor; i.e. the method detects many false positives. Second, the method involves acquiring several high-resolution image volumes before, during and after the injection of a contrast agent. The large volume of data makes the task of interpretation by the radiologist both complex and time-consuming. These shortcomings have motivated the research and development of the computer-aided detection systems designed to improve the efficiency and accuracy of interpretation by the radiologist. Whilst such systems have helped to improve the sensitivity/specificity of interpretation, it is the premise of this thesis that further gains are possible through automated image analysis. However, the automated analysis of breast MRI presents several technical challenges. This thesis investigates several of these, noise filtering, parametric modelling of contrast enhancement, segmentation of suspicious tissue and quantitative characterisation and classification of suspicious lesions.

In relation to noise filtering, a new denoising algorithm for dynamic contrast-enhanced (DCE-MRI) data is presented, called the Dynamic Non-Local Means (DNLM). The DCE-MR image data is inherently contaminated by Rician noise and, additionally, the limited acquisition time per volume and the use of fat-suppression diminishes the signal-to-

noise ratio. The DNLM algorithm, specifically designed for the DCE-MRI, is able to attenuate this noise by exploiting the redundancy of the information between the different temporal volumes, while taking into account the contrast enhancement of the tissue. Empirical results show that the algorithm more effectively attenuates noise in the DCE-MRI data than any of the previously proposed algorithms.

In relation to parametric modelling of contrast enhancement, a new empiric model of contrast enhancement has been developed that is parsimonious in form. The proposed model serves as the basis for the segmentation and feature extraction algorithms presented in the thesis. In contrast to pharmacokinetic models, the proposed model does not rely on measured parameters or constants relating to the type or density of the tissue. It also does not assume a particular relationship between the observed changes in signal intensity and the concentration of the contrast agent. Empirical results demonstrate that the proposed model fits real data better than either the Tofts or Brix models and equally as well as the more complicated Hayton model.

In relation to the automatic segmentation of suspicious lesions, a novel method is presented, based on seeded region growing and merging, using criteria based on both the original image MR values and the fitted parameters of the proposed model of contrast enhancement. Empirical results demonstrate the efficacy of the method, both as a tool to assist the clinician with the task of locating suspicious tissue and for extracting quantitative features.

Finally, in relation to the quantitative characterisation and classification of suspicious lesions, a novel classifier (i.e. a set of features together with a classification method) is presented. Features were extracted from noise-filtered and segmented-image volumes and were based both on well-known features and several new ones (principally, on the proposed model of contrast enhancement). Empirical results, based on routine clinical breast MRI data, show that the resulting classifier performs better than other such classifiers reported in the literature. Therefore, this thesis demonstrates that improvements in both sensitivity and specificity are possible through automated image analysis.

# Keywords

Magnetic Resonance Imaging, Dynamic Contrast Enhanced, MRI, DCE-MRI, breast cancer, denoising, segmentation, lesion classification, medical image analysis

# Australian and New Zealand Standard Research Classifications (ANZSRC)

080106 (Image Processing) 50%, 080104 (Computer Vision) 30%, 080199 (Artificial Intelligence and Image Processing not elsewhere classified) 20%

# Table of contents

# List of figures

# List of tables

# 1. Introduction

Breast cancer is the most common cancer in women and represented over a quarter of all reported cancer cases in Australia in 2006 (Australian Institute of Health and Welfare, 2009). In Australia alone, more than 2500 women die from the disease every year (2007, Australian Institute of Health and Welfare, 2009). Moreover, health expenditure on breast cancer in Australia from 2004 to 2005 was over A$330 million (Australian Institute of Health and Welfare, 2009). The cause of breast cancer is still unknown and thus prevention is impossible. Therefore, early detection and medical imaging-guided therapy form the main strategy for improving mortality rates caused by breast cancer.

Breast cancer screening is commonly based on x-ray mammography owing to its low cost and the short acquisition time that provides a high throughput. X-ray mammography, however, has a high false-negative rate (i.e. low sensitivity) (Behrens et al., 2007, Morris and Liberman, 2005), requiring compression of the breast and is not effective in dense breast tissue (Morris and Liberman, 2005). This has motivated the exploration of alternative imaging modalities including, computed tomography (CT), ultrasound (US), single photon emission computed tomography (SPECT), positron emission tomography (PET) and magnetic resonance imaging (MRI). Of these, MRI shows the most promise for improved breast cancer screening (Morris and Liberman, 2005, Heiberg et al., 1996). In particular, dynamic contrast-enhanced (DCE) MRI shows promise in the characterisation of breast cancer, to which it has a high sensitivity. However, the specificity of the DCE-MRI is relatively low (Heiberg et al., 1996). Also, the acquisition time of the DCE-MRI is longer than X-ray (30–45 minutes as opposed to 5–10 minutes for the X-ray), it is more expensive and requires the injection of a contrast agent (Morris and Liberman, 2005).

## 1.1 Dynamic Contrast-enhanced Breast MRI

Cancers in breast MRI were originally invisible or poorly contrasted (Morris and Liberman, 2005). This has changed with the development of dynamic-contrast-enhanced (DCE) MRI. DCE MRI involves the injection of a contrast agent into the patient that leads to a contrast enhancement over time. However, when the DCE-MRI was first introduced, its technological constraints forced a choice between a high spatial resolution MRI and a high temporal resolution MRI (Orel and Schnall, 2001). As a result, two schools of breast MRI analyses evolved: the dynamic and the static schools.

The dynamic school (mainly in Europe) used a high temporal resolution for characterising suspicious lesions (typically 60 seconds per scan, 5–9 scans per dynamic sequence), while the static school (mainly in the USA) characterised suspicious lesions by evaluating the morphologic features (i.e. shape) at a high spatial resolution and a low temporal resolution (typically, one pre-contrast image and one post-contrast image).

Current technology makes it possible to reach a better compromise between a high spatial and a high temporal resolution. In modern breast MRI, a sequence of three dimensional images of the breast acquired before and during/after the injection of a contrast agent are often used to identify suspicious lesions that are otherwise invisible or poorly contrasted in the non-contrast images (Warren and Coulthard, 2002). The change in signal intensity over time is an important criterion for the differentiation of the malignant from the benign lesions (Warren and Coulthard, 2002). The signal intensity-time curves (enhancement curves) for most cancers show an early steep rise after a contrast agent injection, followed by a plateau and then a washout. Whilst those for benign lesions either show no increase or exhibit a slow continued increase with a delayed washout (Sinha et al., 1997) (see Figure 1.1). Many benign lesions, however, enhance in a similar fashion to cancers (Sinha et al., 1997, Morris and Liberman, 2005) and shallow or non-enhancement is a feature of some malignant changes (Morris and Liberman, 2005). Consequently, although the sensitivity of breast MRI is high (approximately 90%), the specificity of the technique is variable (Jansen et al., 2008, Heiberg et al., 1996) and usually lies between 37% and 86% (Behrens et al., 2007, Heiberg et al., 1996, Jacobs et al., 2003, Liu et al., 1998, Orel and Schnall, 2001, Warren et al., 2005). This has motivated research into ways for improving specificity including, (i) combining the morphologic features of the lesion with its kinetic features and, (ii)

combining the radiologist's interpretation with the quantitative measurements using computer image analysis features (computer-aided detection). Moreover, the subjectivity in the interpretation of breast MRI is high, which has led to the development of a standard method for reporting breast MRI findings.

## 1.2 Standardised reporting of breast MRI findings

The BI-RADS™ (Breast Imaging Reporting and Data System) lexicon for breast MRI (Morris and Liberman, 2005, 2006), published by the American College of Radiology (ACR) provides a standard method for reporting the morphometric and kinetic features on which radiologists should base their analyses. The kinetic features in the lexicon describe the temporal analysis that should be made by examining the enhancement curves averaged over a region of interest within the suspicious lesion (Figure 1.1). This lexicon is aimed at improving the objectivity of the analysis and creating a common language between radiologists for describing the features of suspicious lesions.

## 1.3 Computer assisted evaluation of breast MRI

Dynamic Contrast Enhanced (DCE) MRI data comprises a set of high resolution volumes, acquired before and during/after the injection of a contrast agent. Each volume typically comprises more than a hundred slices (for bilateral acquisition), 256x256 or 512x512 voxels in size (Figure 1.2). In the conventional approach, data are analysed manually by a radiologist. This involves looking for lesions that have certain characteristics, both in the spatial and in the temporal domains.

The amount of data that needs to be interpreted by the radiologist is often huge, and is likely to increase as the spatial resolution of the MRI machines improves with technology. Thus, radiologists can be overwhelmed by the amount of data and the increasing workloads. This motivates the development of computer-assisted evaluation (CAE) systems that will allow radiologists to diagnose the data more efficiently. Furthermore, the specificity of MRI is relatively poor and MR images are difficult to interpret in some circumstances, because of the large amounts of data that need to be analysed in each study.

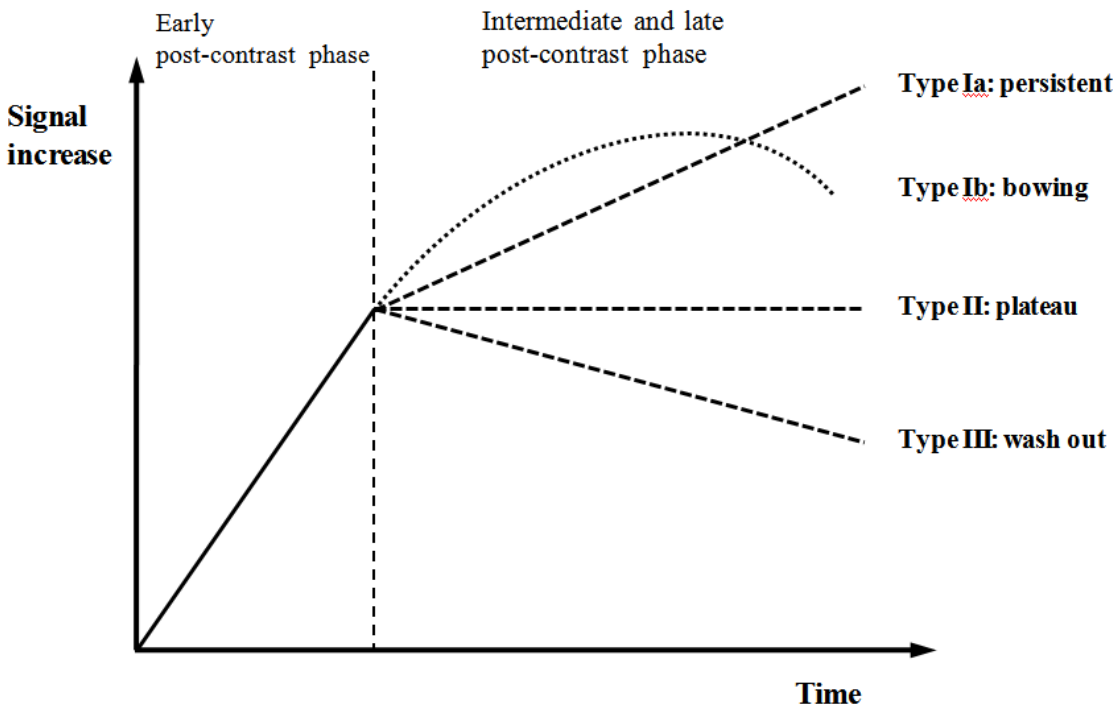



Figure 1.1: Types of enhancement curves in DCE-MRI Adapted with changes from (Kuhl et al., 1999, Morris and Liberman, 2005).

Computer image analysis provides various techniques for medical image analysis. These include, image denoising, registration, segmentation and classification. When dealing with the automated classification of images, a combination of several techniques may be required to achieve a reliable process that will yield consistent results. Also, because the main goal of such techniques is to help the radiologist to perform the interpretation more efficiently, fully-automatic techniques are usually of higher interest than interactive ones.


Figure 1.2: The structure of a typical DCE MRI data set (bilateral axial slices)

## 1.4 Scope of this research

This research focuses on computer-aided analysis techniques. This thesis constitutes part of a larger research project, known as the Breast MRI project, within the Biomedical Engineering Research Division within the School of ITEE at the University of Queensland, Australia. The aim of the Breast MRI project is to improve the specificity and the sensitivity of breast MRI, and therefore its clinical utility, through the use of computer vision (CV), computerised image analysis and pattern recognition techniques and the integration of information concerning breast tissue morphology and contrast-enhancement kinetics. CV offers the possibility of quantitatively, and hence objectively, characterising breast tissue enhancement kinetics and morphology. The underlying hypothesis of this research is that novel quantitative features derived from both dynamic contrast-enhanced (DCE) MRI sequences and co-registered high resolution pre- and post-contrast ($T_1$- and $T_2$-weighted) images will lead to improved sensitivity and specificity, and to more accurate characterisation of lesions. Indeed, recent research suggests that the combination of quantitative features characterising lesion morphology and the features characterising the enhancement curve lead to improved sensitivity and specificity (Chen et al., 2004, Nattkemper et al., 2005, Sinha et al., 1997, Morris and Liberman, 2005).

However, the development of the requisite CV algorithms presents several technical challenges including, noise reduction, correction for patient movement, segmentation of breast volume, bias field correction, parameterisation of contrast enhancement, segmentation of enhancing tissue, feature extraction and the classification of suspicious tissue. This thesis primarily focuses on the following.

> **1. Noise reduction**. Noise in MRI obeys the Rician distribution (Rice, 1944, Macovski, 1996, Gudbjartsson and Patz, 1996). The signal-to-noise ratio (SNR) is especially low when the resolution of the acquired image is high or when the acquisition time is short. Also, fat suppression, which is in wide use, reduces the SNR in the image. A low SNR may mask the fine details in the image and diminish its effective resolution, making both manual and automatic analysis of the image more complicated and less reliable.

**2. Parameterisation of contrast enhancement**. Empiric-parametric and model-parametric (e.g. *pharmacokinetic* models) methods must be devised and investigated. The quality of the model-fitting is critical. For example, in the case of the pharmacokinetic models, they are typically fitted pixel-wise, using iterative non-linear least-squares algorithms, such as the *Levenberg-Marquardt* or the *Nelder-Mead* where fitting can be problematic: slow, fail to converge or extreme parameter values after a fixed number of iterations (Furman-Haran and Degani, 2002, Sykulski et al., 1998, Buckley et al., 1994, Hittmair et al., 1994, Martel, 2006).

**3. Segmentation of enhancing tissue**. A robust algorithm is needed to automatically delineate the regions of clinical significance in each slice or volume.

**4. Feature extraction**. A suitable set of features must be devised for enhancing lesions to construct a feature space that will be used for automatic classification of suspicious lesions.

**5. Classification of enhancing tissue**. Based on the most discriminatory features, a classifier should be developed that can discriminate benign and malignant lesions.

## 1.5 Research hypothesis

Recent research shows that combining morphometric and kinetic information has the potential for improving the accuracy of interpretation of DCE-MRI of the breast (Chen et al., 2004, Nattkemper et al., 2005, Sinha et al., 1997, Morris and Liberman, 2005, Szabo et al., 2003). Nevertheless, the specificity of DCE-MRI is still too low and breast MRI interpretation is usually based solely on the radiologist's analysis. Hence, **the hypothesis of this research is that the specificity (and possibly the sensitivity) of breast MRI interpretation can be improved** by:

- Designing features for suspicious lesion classification that will integrate information about tissue kinetic enhancement and morphology

- Reducing the subjectivity of breast MRI interpretation by using image analysis and pattern recognition techniques to automatically classify suspicious lesions in the breast.

## 1.6 Aims and objectives

To test this hypothesis, this research focuses on the development and evaluation of CAE system that will automatically classify suspicious lesions in breast MRI. This system will be constructed using 'low-', 'intermediate-' and 'high-level' processing methods (Gonzalez and Woods, 2002). It will use novel image processing and computer vision algorithms and will improve the detection and characterisation (differentiation, size, extent) of breast cancer in MR images. To this end, it was necessary to undertake the following tasks.

1. **The development of a new denoising method for improving the quality of DCE MR Images.** Because of the nature of the MRI systems, DCE MRI data is contaminated by Rician noise. The noise in the images may often mask fine details and diminish the effective resolution. To perform a high level analysis on the image, it is essential that the input data have the least possible amount of noise. In this research project, a new denoising algorithm for DCE-MRI will be developed.
2. **The development of a new model of contrast enhancement to be used as a tool for automatic/semi-automatic/visual analysis of DCE MRI data.** Models of enhancement for DCE MRI help by reducing the dimensionality of the data from typically 5 to 9 volumes to as little as three free parameters of the model that can then be used for automatic analysis. The models of enhancement also mitigate the effect of noise in the data. In this research project, a model of enhancement that has no more three free parameters will be developed and evaluated against existing models.
3. **The development of a robust segmentation method for suspicious lesions in the breast.** Automatic segmentation of medical data is challenging because of the delicate nature of the problem (low tolerance to misinterpretation). A reliable segmentation method is highly desirable, especially one that has almost no false negatives and a minimum amount of false positives. As part of this

research, a new algorithm for automatic segmentation of enhancing lesions in DCE-MRI of the breast will be developed.

4. **The design and implementation of a feature space and a classifier for automatic classification of suspicious lesions in the breast.** The interpretation of breast MRI images is usually performed in both the spatial and the temporal domains. The BI-RADS lexicon, in addition to previous work describing automatic classification of suspicious lesions in breast MRI, is used as a basis for the construction of an efficient feature space. Classifying suspicious lesions with a high precision requires an efficient feature space and also a suitable classifier. A well-designed classification system may thus reduce the subjectivity in diagnosis and improve the specificity of the DCE-MRI. As part of this research, a feature space and a suitable classifier for the classification of suspicious lesions in the DCE-MRI of the breast will be developed.

## 1.7 Overview of the thesis

The remainder of the thesis is organised as follows.

Chapter 2 intends to familiarize the reader with some basic terms and concepts in MRI, breast cancer and statistical pattern recognition that will later be used in different places in this thesis. It introduces the terminology and background concepts of MR Imaging of the breast and breast cancer. It describes the basic physics behind different types of MR imaging in addition to the physiology described. The chapter then reviews the different types of breast cancer and how they appear in MR imaging. Finally, the chapter provides an overview of some basic concepts in the field of statistical pattern recognition that are relevant for this thesis.

Chapter 3 reviews existing models for contrast enhancement in DCE-MRI, including a new, empirical, model of enhancement. In addition, the chapter presents a comparison between the 'goodness of fit' of the existing models and the new one. The chapter also describes the properties of the new model and evaluates it, in comparison with the existing pharmacokinetic models.

Chapter 4 describes the noise model in MRI and various denoising methods. The chapter reviews denoising methods for MRI and describes a new denoising method for DCE-MRI; the Matlab code for this method is given in Appendix C. This paves the way for

the next chapters, where the high performance of the proposed methods relies on noise-free DCE-MRI data.

Chapter 5 presents a method for the automatic segmentation of enhancing lesions in DCE-MRI of the breast, which is considered an 'intermediate level' processing. The method is based on entropy-based binarization (described in Appendix A) and a seeded region growing (described in Appendix B) where the selection of seeds is performed automatically, based on a set of criteria. This method can be used as a tool for reducing the volume of data that needs to be interpreted by the radiologist or as a basis for automated classification of suspicious lesions.

Chapters 3–5 describe the novel CAE system that was developed during this research. The methods that incorporate this system can be classified into three groups, 'low-level', 'intermediate-level' and 'high-level' image analysis methods. Chapters 3–4 of the thesis describe 'low level' methods for improving or enhancing DCE-MR images, including image denoising and kinetic modelling.

Chapter 6 describes a 'high-level' processing method that uses the concepts and methods that were previously described in this thesis. The chapter describes a selected set of features and a classifier for automatic classification of suspicious lesions in DCE-MRI of the breast. Also, the chapter describes an evaluation of the classification of suspicious lesions in real data, using the selected features and classifier.

Chapter 7 summarises the main points arising from this research and draws some conclusions from them. Also, it describes possible future directions in which this research may continue.

# 2. Background

This research draws upon several fields, including: mathematics, image processing, computer vision, pattern recognition and biomedical engineering. In Chapter 1, breast MRI was introduced and its advantages and disadvantages relative to other imaging modalities were discussed. In this chapter, a general background in a variety of relevant topics is provided. The first part of the chapter describes the basic principles in MRI physics. It then describes the anatomy of the human breast and breast mammography. This material is required for understanding the material in the later chapters of this thesis. The last section of the chapter provides a basic background of statistical pattern recognition, including basic classifier types and methods for evaluating classification performance, which is needed for understanding the methods that are described in Chapter 6.

## 2.1 The theory of MRI

This section acquaints the reader with the underlying physics of MRI and of its clinical application, particularly to DCE MRI of the breast. This section provides the reader with basic knowledge of the imaging method (DCE-MRI) that is later used as the primary source of clinical data for this research. The majority of the material in this section is primarily based on (Warren and Coulthard, 2002, Haacke et al., 1999).

### 2.1.1 Basic physics of MRI

Subatomic particles, such as protons, have the quantum property of spin. Magnetic Resonance signals are a result of the interaction between a magnetic field and the spin angular momentum of atomic nuclei (Leggett, 2004). More specifically, the precession of the hydrogen protons yields changes in the flux in the nearby coils. These changes in flux are used to create an MR image. Hydrogen nuclei are most commonly used, mainly because they are more abundant in the human body than any other nucleus capable of

undergoing nuclear magnetic resonance (NMR) and thus give a denser signal for a given period of time.

In the absence of a magnetic field, nuclear spins do not have any preferred direction of alignment. Each nucleus spins around an axis called the magnetic moment. Once placed in a strong magnetic field, $B_{0,}$ producing bulk (averaged) nuclear magnetization, $M$, the magnetic moments of the bulk magnetisation (i.e. the average magnetic moment) will tend to align with the direction of the magnetic field (Figure 2.1).

The magnetization, $M$, has a naturally-preferred alignment in the direction of $B_0$. Nuclei with such a property are called nuclear spins, where the Z axis denotes the initial alignment of the top. The bulk magnetisation of the protons is then tipped away from the external field direction to produce a magnetic field that yields changes in the flux in any nearby coil (Haacke et al., 1999).

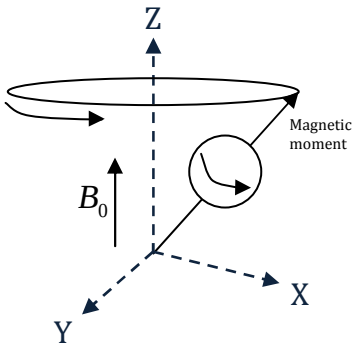



Figure 2.1: Larmor precessing of a nuclear spin around an applied magnetic field, $B_0$

In NMR the precession frequency $\omega_0$ is called the Larmor frequency and is given by:

$$\omega_0 = \gamma \cdot B_0, \tag{2.1}$$

where $\gamma$ is a scalar called the gyromagnetic ratio and is measured in radians per second per Tesla (unit of magnetic flux density). The values of $\gamma$ are different for different nuclei. From equation (2.1) it can be seen that the precession frequency is directly proportional to the static field strength. A typical field strength used in clinical MRI is 1.5 Tesla in which the Larmor frequency for hydrogen is 63.9 MHz.

To measure the bulk magnetization, $M$ has to be disturbed by an oscillating magnetic field applied at right angles to $B_0$. The field has to be applied at precisely the Larmor frequency to produce the effect because resonant absorption of energy by the protons

due to an external oscillating magnetic field occurring exactly at the Larmor frequency. The oscillating field is usually referred to as the Radio Frequency (RF) magnetic pulse, because it is applied for only a few milliseconds. An RF pulse that rotates *M* through 90° from its initial position aligned with Z is called a 90° pulse. If the amplitude of the RF pulse is doubled or, alternatively, if it is applied for twice as long, then *M* is rotated by 180°. The strength of the magnetic field, $B_1$, of the RF pulse is typically in the order of $10^{-5}$ $B_0$.

After the nuclear magnetization *M* is moved away from its initial alignment by an RF pulse, it will begin to realign itself as soon as the RF pulse is switched off. The *z* component of the magnetization $M_z$ recovers exponentially with time constant $T_1$, toward its equilibrium value $M_0$, the value at which *M* is aligned with $B_0$. The MR signal is produced by the transverse magnetisation of the precessing spins in the measured volume of the body. This signal decays in amplitude and is detected externally, often by the same coil that produces the RF pulse (Poole, 2007). $T_1$ is called the longitudinal relaxation time or spin-lattice relaxation time. The actual meaning of $T_1$ is that the difference between $M_z$ and $M_0$ is decreased by 63% of its value in each $T_1$ period, provided that no additional RF pulses are applied. The relaxation process can be described as follows:

$$M_z(t) - M_0 = (M_z(0) - M_0) \cdot e^{-\frac{t}{T_1}}, \tag{2.2}$$

where $M_z(t)$ is the longitudinal magnetisation (in the direction of $B_0$) and $M_0$ is its equilibrium value.

In a similar way, the amount of any magnetization rotated into the transverse plane, $M_{xy}$, declines during the recovery to equilibrium. As with longitudinal relaxation, the decay to the final value of $M_{xy} = 0$ is exponential. This process can be described as follows:

$$M_{xy}(t) = M_{xy}(0) \cdot e^{-\frac{t}{T_2}}, \tag{2.3}$$

where $M_{xy}(t)$ is the transverse component of the magnetisation and $T_2$ is the time constant of the exponential decay of the transverse magnetisation. It is often called 'transverse relaxation time' or the 'spin-spin' relaxation time.

$T_2$ is always less than or equal to $T_1$. The meaning of $T_2$ is that $M_{xy}$ decreases by 63% of its value in each $T_2$ period in the absence of any RF pulses. MR imaging has been developed to show the differences in these relaxation time constants.

### 2.1.2 **Encoding the MR signal**

The underlying principal in MR imaging is that the Larmor precession frequency is used to mark the position of the encoded volume. The precession frequency, $\omega_0$, of an NMR signal is directly proportional to the strength of the static magnetic field $B_0$ (Equation 2.1). A magnetic field gradient coil can change the strength of $B_0$ as a function of the position within the scanner. Thus, the Larmor frequency varies along the direction of the gradient coil (Figure 2.2).

Given that only one dimension can be encoded at a time, three separate gradient magnetic coils are used in an MRI scanner, $G_x$, $G_y$ and $G_z$, one for each axis. The imaging process therefore encodes each one of the directions sequentially by alternately turning on each one of the gradient fields.

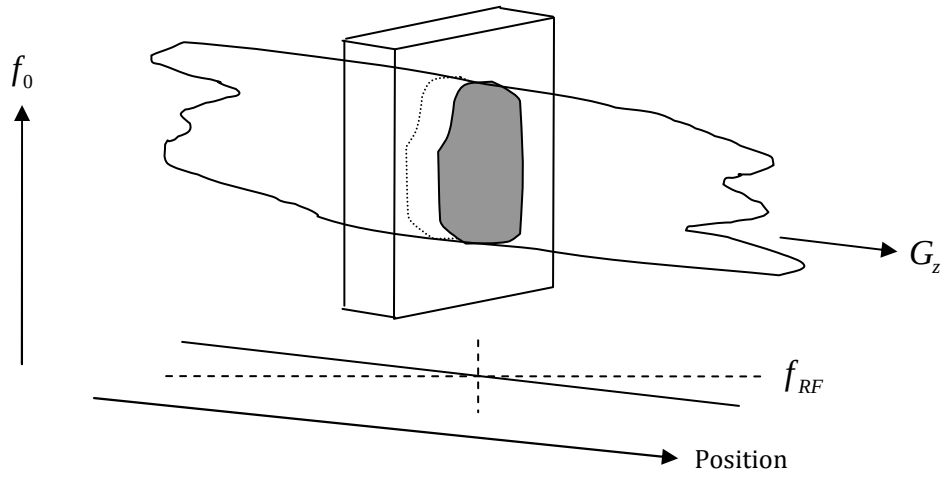



Figure 2.2: The magnetic fields in a single slice MR imaging
A magnetic field gradient, $G_z$, is applied at the same time of an 90° RF pulse. The frequency of the RF pulse, $f_{RF}$, matches exactly the Larmor frequency, $f_0$, at the position of the imaged slice.

Once the gradient has been switched on, a 90° (flip angle) RF pulse is applied. A 90° RF pulse is considered to be the simplest and the signal resulting from it is called the free induction decay (FID). Nuclei with a Larmor frequency that matches the RF pulse frequency are then rotated through 90° and precess to yield an NMR signal. The position

of the selected slice is changed by changing the frequency of the RF pulse. This frequency depends on the position along the *z* axis:

$$\omega = \gamma \cdot (B_0 + G_Z \cdot z) \tag{2.4}$$

To excite a slice, a range of frequencies $\omega_1 < \omega < \omega_2$ have to be produced, corresponding to the range of frequencies inside the excited slice. To do that, an RF pulse is used. The RF pulse's shape is selected to be a sinc shape that creates a 'top hat' shape response that is 1 inside the frequency interval of the slice, and 0 outside of this interval. This happens because the Fourier transform of the sinc function is a rectangular function.

The slice thickness is typically a few millimetres and is controlled by the frequency spread contained within the 90° RF pulse.

To spatially encode the NMR signal, a frequency-encoding gradient is switched on at the moment the slice selection gradient has been switched off. Protons at different positions along the frequency encoding direction will therefore precess at different frequencies and will generate different frequency signals in the receiving coil. To separate the signals and create a readable image, an inverse Fourier transform is then performed.

The raw data, received by the RF receiver coil of a MR scan is known as k-space. The inverse 2D Fourier transform of the k-space is the MR image that will be interpreted by a clinician.

### 2.1.3 **Pulse sequences**

Various pulse sequences are used in MRI. Two of the most common ones are the spin echo (SE) and gradient echo (GE) pulse sequences. The GE is especially important for dynamic contrast-enhanced (DCE) MRI acquisition.

In the case of an SE sequence, a 90° RF pulse is applied. After half of the echo time (TE) has passed, a 180° RF pulse is applied, which re-phases the precessing spins at time TE. This pulse compensates for signal loss caused by inhomogeneities in the magnetic field and maximises the signal emission. The SE sequence will result at TE. The whole sequence is repeated with a repetition time (TR), depending on the desired image (i.e. $T_1$-weighted, $T_2$-weighted or proton-weighted).

In the case of GE sequences the echo is produced by reversing the gradient field, which causes a re-phased RF echo (i.e. gradient echo). In addition, the initial 90° RF pulse may

be substituted with a pulse with a (RF pulse) flip angle smaller than 90°, which does not use the entire longitudinal magnetisation. This allows faster image acquisition at the expense of decreased signal intensity. GE sequences suffer from higher susceptibility to image artefacts than SE sequences owing to a greater sensitivity to field inhomogeneities (Fischer and Brinck, 2004).

### 2.1.4 **2D and 3D imaging**

In 2D MR imaging, a single slice is excited at a time. The slices are adjacent and, ideally, have no gaps between them. The field of view (FOV) of each slice is rectangular and approximately 2–3 mm in thickness. In 3D MR imaging, the entire breast is excited as a volume. The TR values for 2D imaging are usually between 200 and 300 ms. The ideal flip angle is between 70° and 90°.

The 3D acquisition technique allows thin slices to be acquired with no gaps (typically 2 mm thick). Another advantage of the 3D acquisition technique is a shorter acquisition time. This is the result of the shorter repetition time (TR) that is usually in the order of 10 ms. The flip angle for 3D imaging is typically 25°. However, 3D MRI acquisition suffers from a higher susceptibility to artefacts and requires a higher dose of contrast agent (see Section 2.1.8 for further details about contrast agents) (Fischer and Brinck, 2004).

### 2.1.5 **Tissue contrast in MRI**

Although most of the tissues in the human body have a similar water or proton density, in MRI the NMR signal strength is greatly influenced by the $T_1$ and $T_2$ relaxation times. This, in turn, influences the intensity of the different tissues in the displayed image. Disease can also considerably alter the signal strength of a tissue.

The longest $T_1$ relaxation times usually appear in body fluids. Body fluids also have long $T_2$ values. However, relaxation times are greatly decreased by the presence of blood. In practice, the differences between $T_1$ and $T_2$ are used to produce $T_1$ and $T_2$-weighted images (Warren and Coulthard, 2002), as demonstrated in Table 2.1. By altering the acquisition parameters, different tissue types can be highlighted in the image.

Table 2.1: Proton $T_1$ relaxation times for some types of tissue at 0.5, 1.0 and 1.5 $T$
$T_2$ does not vary greatly with Larmor frequency (Warren and Coulthard, 2002).

| **Tissue** | $\boldsymbol{T_1(0.5)T}$ **ms** | $\boldsymbol{T_1(1.0)T}$ **ms** | $\boldsymbol{T_1(1.5)T}$ **ms** | $\boldsymbol{T_2}$ **ms** |
|---|---|---|---|---|
| **Grey matter** | - | 1040 | 1140 | 100 |
| **White matter** | 450 | 660 | 720 | 90 |
| **Muscle** | 560 | - | 1160 | 35 |
| **Cerebral Spinal Fluid (CSF)** | 4000 | 4000 | 4000 | 2000 |
| **Liver** | 360 | - | 720 | 60 |

### 2.1.6 **Bias Field**

The bias field is a phenomenon that causes the same type tissue in different locations in the MR image to have different levels of intensity. It is identified as a low spatial frequency signal that increases the average intensity in some parts of the image and reduces it in others. The bias field originates in non-uniformity in the $B_1$ excitation field owing to non-uniformity in the interaction between the RF field and the tissue of the patient being imaged. This non-uniformity results in different amounts of signal being received from tissue in different spatial locations (Hayton, 1998).

### 2.1.7 **Fat and Silicone suppression**

Fat usually has high intensity on $T_1$- and $T_2$-weighted images unless suppressed. Bright fat can sometimes obscure adjacent tissue and can introduce artefacts. Two main methods are used to suppress the fat signal:

1. The short Tau Inversion Recovery (STIR) sequence
2. Chemical shift saturation

In the first method, a 180° RF pulse is used to invert the spin alignment from $Z^+$ to $Z^-$. No signal is produced by this process because no magnetization is introduced in the transverse plane. At a time ($TI$) later, a 90° RF pulse, or a (90°, 180°) RF pulse pair, is used to tip the magnetization into the transverse plane to generate a gradient-echo or spin-echo signal, respectively. The amount of z-magnetization ($M_Z$) presented immediately before the 90° pulse determines the signal obtained. The $T_1$ of fat is shorter than most other tissues. Therefore, $TI$ is chosen such that the 90° RF pulse is applied at exactly the time that the fat $M_Z$ has recovered to zero (Figure 2.3). Note that other tissues with longer $T_1$ values will have a negative $M_Z$ value at that point. However, in the resulting image, only the magnitude of $M_Z$ is important and the sign is ignored.

In the second method, the fat signal is suppressed, based on the small difference in the Larmor frequency of fat protons compared to water protons. The difference in the resonant frequency $\omega_0$ is called a chemical shift, because it arises from the different magnetic environment of the hydrogen protons. It is therefore possible to apply a 90° RF pulse tuned to the fat molecules, but not to the water molecules, provided that the field gradients have been switched off. The z-magnetization of the fat molecules will be zero after the pulse, thus it can be followed by a conventional gradient or spin echo sequence, which will yield a very low fat signal and thus a low SNR that results in a 'fat free' noisy image.

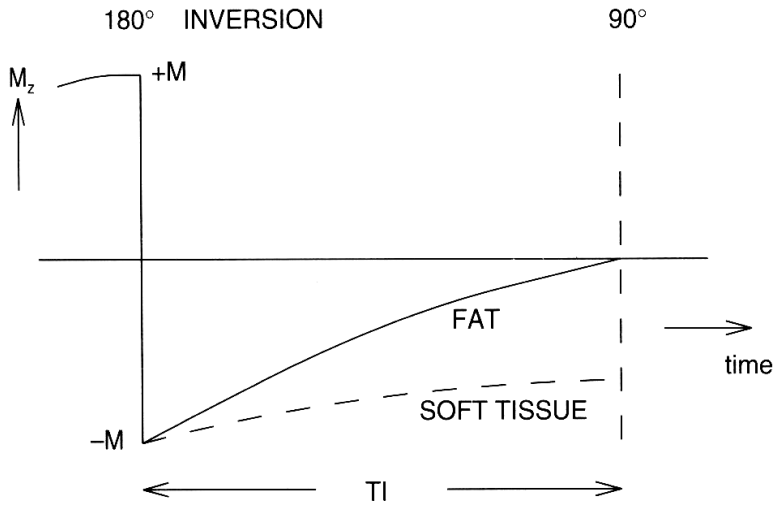



Figure 2.3: STIR sequence fat suppression
TI is chosen such that the magnetization in the Z direction is zero for fat, but not zero for most other tissues. Thus fat will yield no signal in the image (reproduced from (Warren and Coulthard, 2002)).

When imaging breast implants, the silicone filling material produces a high signal, which can mask enhancing tissue. However, because silicone protons are chemically shifted from water and fat protons, a selective saturation process can be applied, just as in the chemical shift suppression of fat.

### 2.1.8 **Contrast agents and dynamic studies in Breast MRI**

MR breast imaging often uses $T_1$ and $T_2$-weighted images pre- and post-contrast (i.e. before and after the injection of a contrast agent into the bloodstream) to identify suspicious lesions, because many cannot be detected in conventional $T_1/T_2$-weighted

(non-contrast) images. Malignant lesions, as well as some benign conditions, show contrast enhancement after the injection of a contrast agent. The contrast enhancement is primarily due to angiogenesis (growth of new blood vessels) around the malignant tissue, which accelerates the blood inflow, and hence the flow of contrast agent around the tissue. Also, malignancy-related angiogenesis creates vessels with 'leaky endothelial linings' (Morris and Liberman, 2005), which increase the flow of contrast agent in the extracellular compartment at the site of the tumor. This allows the creation of contrast-enhanced MR images that are created by subtracting the pre-contrast from each post-contrast image and thereby providing a better contrast for malignant lesions.

Most malignant tissues enhance in contrast-enhanced MRI (i.e. 1 pre-contrast, 1 post-contrast), which makes it sensitive to breast cancer (Morris and Liberman, 2005, Warren and Coulthard, 2002). However, it can be difficult to distinguish benign from malignant disease, because some benign conditions also exhibit contrast enhancement. This limitation can be diminished by using dynamic contrast-enhanced (DCE) MRI. DCE MRI involves taking a series of sequential $T_1$-weighted images every few seconds, or tens of seconds (typically 60–90), following a bolus injection of Gd-DTPA (gadolinium-diethylene-triamine pentaacetic acid; gadopentetate dimeglumine). The result of the acquisition is a series of volumes resulting from the $T_1$ acquisitions at the different time points, as demonstrated in Figure 1.2. In this case, both the rate of signal change in addition to the characteristic shape of the signal versus time is used to interpret the image and identify suspicious lesions (Figure 1.1).

### 2.1.9 **The relationship between signal enhancement and contrast agent perfusion**

The characteristic behaviour of enhancement curves in DCE-MRI is related to excessive angiogenesis, the growth of new blood vessels, around many types of malignant tissues. Angiogenesis is a natural process, occurring both in healthy and diseased tissues in the body. In some malignant conditions, the body loses control of the angiogenesis process and excessive angiogenesis develops. Tumours cannot enlarge beyond 1 to 2 mm unless they are vascularised; thus, angiogenesis is a requisite for continued tumour growth, in addition to metastasis (secondary growth of malignant tissues around the body). Hence, angiogenesis is a necessary biologic condition of malignancy (and some benign disease as well) (Morris and Liberman, 2005).

In DCE MRI, angiogenesis facilitates contrast enhancement in two ways, increased vascularity leads to an increased contrast agent inflow, and increased vessel permeability leads to an accelerated contrast extravasation at the tumour site (Morris and Liberman, 2005), which enhances the exchange of the contrast agent between the tissue compartments. The molecular weight of the Gd-DTPA contrast agent allows it to diffuse outside the blood vessels into the extra-cellular compartment, but not to penetrate the cell membrane. Figure 2.4 shows the major tissue compartments involved in the distribution of contrast agent (represented by stars).

### 2.1.10 **Interpreting DCE-MR images**

Contrast enhancement in DCE-MRI is related both to malignant and benign disease. However, contrast enhancement is also related to some healthy tissue such as the liver. To properly interpret DCE-MRI and to be able differentiate between malignant and benign lesions, a deep knowledge of breast and human anatomy is required. This knowledge provides contextual information that can easily be used by humans, but, to date, can hardly be automated for use in CAE tools.

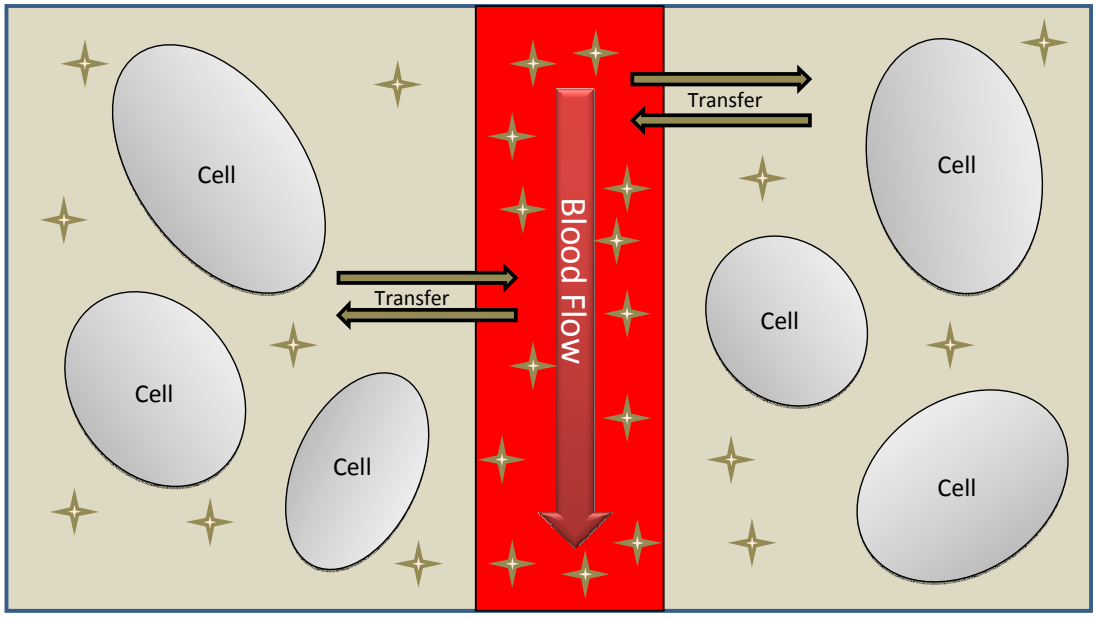



Figure 2.4: Major compartments and functional variables involved in the distribution of a contrast agent

## 2.2 **Breast MRI**

This section focuses on breast MRI and its interpretation. It describes the anatomy of the human breast, the types of breast cancer and their characteristics of appearance in breast MRI.

### 2.2.1 **Anatomy of the human breast**

The human breast is actually a skin gland, enveloped in a fibrous fascia (Morris and Liberman, 2005). The breast content is bounded by the skin on the outside and by the pectoralis major muscle on the back side (which marks the beginning of the chest wall). There are many layers between the breast and the pectoralis major muscle. However, the breast is not completely separated from the pectoralis major muscle and there are lymphatic and blood vessels that penetrate the breast. Breast tissue is divided into parenchyma (glandular tissue) and stroma (connective/fibrous tissue). The parenchyma consists of 15 to 20 lobes (milk glands) that converge toward the nipple. Ducts from the lobes converge into 6 to 10 major ducts that hold a ductal ampulla, beneath the nipple and connect to the outside through the nipple (Figure 2.5). The lobes are arranged in segments of glandular tissue that are connected by stromal tissue (Morris and Liberman, 2005). A segment of lobular tissue, connected to a duct, is called Terminal Duct Lobular Units or TDLU (Figure 2.5). The stromal tissue is mainly fatty tissue and ligaments that surround the lobes and ducts in the breast.

Breast cancer can develop in each of the breast tissue types. Different tissue types may develop different types of cancer that may have different characteristics in DCE-MR images, thus making the description of malignant lesions more complicated.

### 2.2.2 **Types of breast cancer**

Breast cancer is a heterogeneous disease which has several subtypes (Claus et al., 1993). In general, breast cancer can be divided into two types:

1. Carcinoma in situ – when the malignant mass stays confined inside the tissue in which it has developed
2. Invasive carcinoma – when the malignant mass invades surrounding tissue.

Several histological subtypes of breast cancer are known. Of these, the most common ones are (Claus et al., 1993):

1. Invasive ductal carcinoma (IDC). Starts in a duct, then breaks through the basement membrane (i.e. the wall of the duct) and invades the stromal tissue.
2. Ductal carcinoma in situ (DCIS). Cancerous cells develop inside a duct, but do not penetrate the basement membrane.

3. Invasive lobular carcinoma (ILC). Starts in a lobular gland and invades the surrounding tissue.
4. Lobular carcinoma in situ (LCIS). Also called lobular neoplasia, begins in a lobular gland, but does not penetrate the gland's wall.
5. Medullary carcinoma. This is an invasive breast cancer that has a well-defined boundary between the cancerous tissue and the surrounding tissue.

Some of the less common breast cancer types include the colloid carcinoma, tubular carcinoma and adenoid cystic carcinoma.

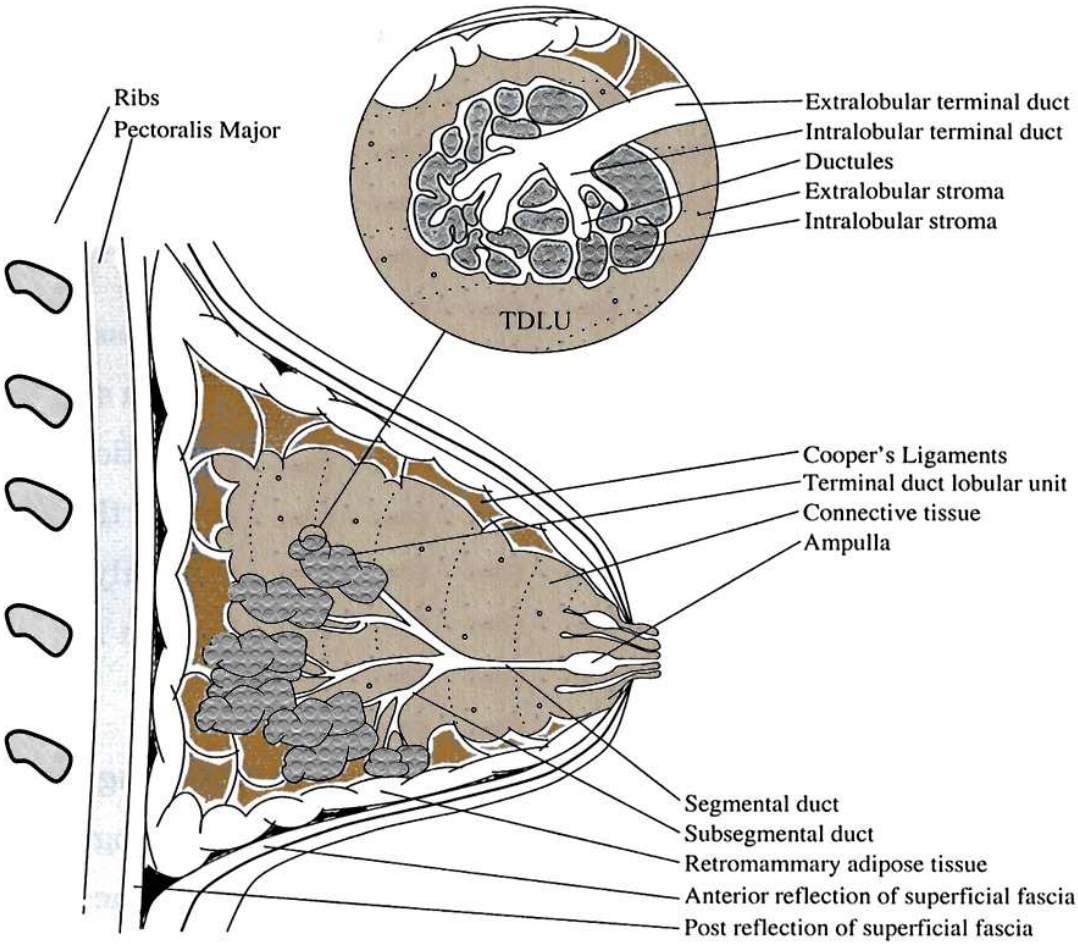



Figure 2.5: The anatomy of human breast
Adapted with changes from (Hayton, 1998).

In DCE-MRI, ductal carcinoma tends to show a linear enhancement similar to that of a blood vessel. Mass enhancement, on the other hand, is usually easier to spot and analyse if the lesion size is sufficient. Improving the specificity of DCE-MRI may thus improve

both detection and categorisation of breast cancer and may help differentiate between enhancing normal tissue (e.g. blood vessels) and enhancing malignant tissue (e.g. DCIS).

### 2.2.3 **Breast MRI mammography**

The prevention of breast cancer is still impossible and thus the main treatment strategy relies on early detection, using mammography. Owing to its relatively high cost, DCE-MRI of the breast is usually reserved for cases with high probability or known malignancy and for cases where other imaging techniques (e.g. ultrasound, mammography) cannot provide a definitive answer. In most cases, X-ray mammography is performed first and only high risk patients are referred to DCE-MRI (Morris and Liberman, 2005). Other imaging methods include Computed Tomography (CT), Single Photon Emission Tomography (SPECT), Positron Emission Tomography (PET) and Tomosynthesis. These include a relatively high level of exposure to ionizing radiation, both on the patient's side and on the technologist's side. In SPECT and PET, a radio-isotope is injected into the patient blood stream and the photons that are emitted from the patient's body, during the radioactive decay process, are then received by a detector to create the image.

#### *Characteristic appearance of benign and malignant breast diseases in MR images*

In $T_1$-weighted images, both normal breast tissue and fibrous tissue (i.e. connective tissue that is not muscles) show a low signal intensity and fat shows an intermediate to high signal intensity. Most benign and malignant lesions also show a low signal intensity on $T_1$-weighted sequences and cannot be differentiated from normal breast tissue on non-enhanced $T_1$-weighted images. In $T_2$-weighted images, fat is of an intermediate signal intensity. The signal intensity of the breast tissue depends on the water content, varying from a low signal intensity in fibrosis to a high or very high signal intensity in the majority of cysts. In contrast-enhanced images, normal breast tissue demonstrates only a slight increase in signal intensity, with some exceptions (such as blood vessels). Malignant lesions enhance, but there are also benign lesions that may enhance in a similar fashion.

Patients with benign breast changes may show delayed and diffuse patchy enhancement in 25–30% of cases (Figure 2.6). However, in 5–10% of cases, there may be focal enhancement, which may be rapid and simulate malignancy (Figure 2.7).

Although tumours can be identified within fatty tissue, the differentiation of benign and malignant tumours cannot be undertaken with certainty using signal characteristics on $T_1$-weighted or $T_2$-weighted sequences, except in the case of a cyst. The use of intravenous gadolinium has increased both the sensitivity and specificity of breast MRI, because most malignant tumours enhance markedly (Warren and Coulthard, 2002).

Following the injection of a contrast agent, most cancers show an early steep rise in enhancement within the first 5 minutes, typically by 70–100% (a threshold that gives high sensitivity, but low specificity). If a lesion enhances by less than 60% or does not enhance at all, it is most likely to be benign, although up to 10% of cancers will also enhance slowly (Warren and Coulthard, 2002).

Time/intensity enhancement curves have been studied by many, including (Hayton et al., 1997, Furman-Haran and Degani, 2002), in an attempt to improve specificity without reducing sensitivity. The shape of the intensity curve was assessed and three types were defined. All three curves demonstrate a rapid increase in signal intensity in the early post-contrast phase. The difference is in the intermediate and late post-contrast phase (Figure 1.1). The prediction accuracy of the time/intensity curves is demonstrated in Table 2.2.

Table 2.2: Percent of benign and malignant cases for each type of curve
Curves are shown in Figure 1.1. Figures in the table are taken from (Morris and Liberman, 2005, Warren and Coulthard, 2002)

| Curve Type | % of benign | **% of malignant** |
|---|---|---|
| Type I (a & b) | 84% | **9%** |
| Type II | 11.5% | **34%** |
| Type III | 5.5% | **57%** |

The pattern of contrast enhancement may also be helpful in differentiating benign from malignant lesions, because malignant lesions show a peripheral enhancement with centripetal progression, whereas enhancing benign lesions may either show a peripheral enhancement with no progression or may enhance centrally rather than peripherally.

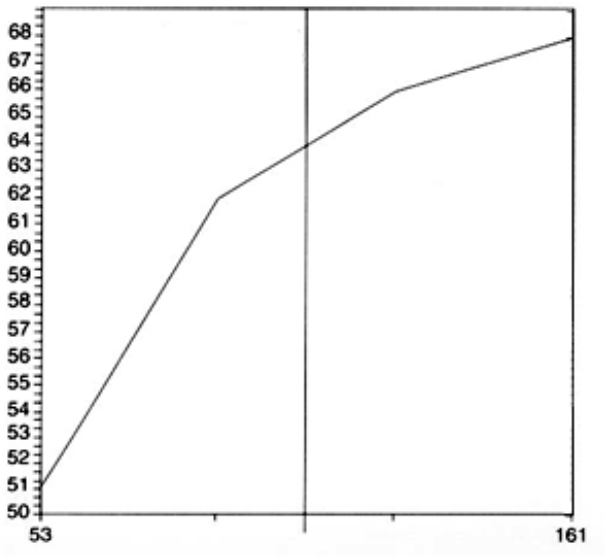



Figure 2.6: The enhancement over time of benign breast tissue (25–30% of the cases)
The vertical axis represents signal enhancement while the horizontal axis represents time. Reproduced from (Warren and Coulthard, 2002).

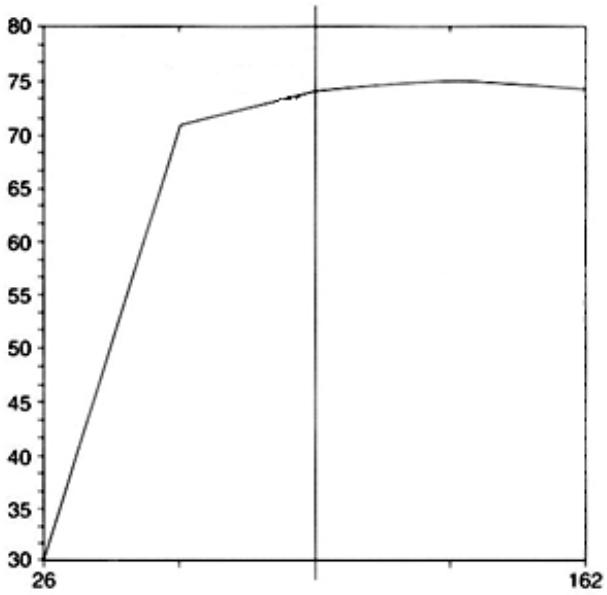



Figure 2.7: The enhancement over time of benign breast tissue which mimics malignancy (5–10% percent of the cases)
The vertical axis represents signal enhancement while the horizontal axis represents time. Reproduced from (Warren and Coulthard, 2002).

Collectively, the amount and speed of the enhancement and the morphological appearances of the lesion have to be considered. Positive enhancement will occur with several benign breast diseases.

### *Architectural features of benign and malignant lesions*

As the use of DCE MRI becomes more common, standardisation of the terminology has become necessary. The need for a lexicon becomes apparent when interpreting contrast-enhanced breast MRI, because there are many different types of enhancements seen in addition to different types of tissue architectures. To reduce the subjectivity of interpretation an architectural lexicon, describing both the morphology and the kinetic behaviour of lesions has been developed.

The American College of Radiologists (ACR) BI-RADS™ (Breast Imaging Reporting Data System) MRI lexicon, first published in 2003, provides a terminology for describing both the enhancement behaviour and the morphology of a lesion (see Appendix D). The ACR lexicon was introduced to help improve the specificity of breast MRI by providing standardised and universally-acceptable terminology for the reporting of breast MRI findings. Morris and Liberman (Morris and Liberman, 2005) present an architectural features lexicon that was assembled over a few years by an international working group consisting of breast MRI experts from around the world and is also based on the BI-RADS™ lexicon. The lexicon describes a list of terms that are usually related to suspicious tissues in breast MRI. By using the lexicon, a common language for describing suspicious tissues can be created that will lead, in turn, to a better diagnosis and cooperation between radiologists. The lexicon that Morris and Liberman present describes the following terms.

*Description of Terms*

Focus/Mass/Stippled **-** A focus is a single tiny punctate enhancement that is non-specific and too small to be characterized. An enhancing lesion on MRI can be described as a mass if it displaces tissue and has space-occupying properties. If there are multiple foci in the breast, the term stippled can be applied.

Shape/Margin **-** Mass shape can be described as round, oval, lobular or irregular. Margins of masses can be described as smooth, irregular and spiculated.

Spiculated and irregular masses are suspicious for carcinoma; whereas, a smooth margin is more suggestive of a benign lesion.

In general, margin and shape analysis should be performed on the first post-contrast image to avoid washout and progressive enhancement of the surrounding breast tissue.

Internal enhancement – Internal enhancement of lesions can be divided into homogenous and heterogeneous. Homogenous enhancement is uniform. Heterogeneous enhancement is non-uniform, with areas of variable signal intensity. Masses may display rim enhancement, a particularly suspicious finding for malignancy. Other suspicious findings include enhancing septations or central enhancement, though these signs occur less commonly.

Homogeneous enhancement is suggestive of a benign process. Other benign lesions include an inflammatory cyst that enhances peripherally and benign fat necrosis that can exhibit rim enhancement with a central low signal indicating fatty content.

Nonmass enhancement – If the enhancement is neither a focus nor a mass, then it is classified as non-mass-like enhancement. A non-mass enhancement is classified according to the distribution of the enhancement and can be described as linear-ductal, linear-nonspecific, regional, segmental or diffuse. Linear enhancement is usually related to the ductal system. Ductal enhancement corresponds to one or more ducts in orientation and is suspicious for DCIS (ductal carcinoma in situ). Ductal non-specific will not follow this pattern and is less suspicious for malignancy.

Segmental refers to an enhancement that is triangular in shape with the apex at the nipple and is suspicious for DCIS with a single branching duct system.

Regional enhancement is an enhancement that does not correspond to a single duct system and can be further described as smooth or linear smooth (suggestive of benign) or irregular, which is a non-smooth enhancement, continuous or discontinuous (suspicious for malignancy, especially if clumped).

Regional, segmental or diffuse enhancement can be further described as homogeneous, heterogeneous-stippled/punctate (scattered multiple foci, 1–2 mm), clumped, septal or non-specific. These are usually demonstrated by benign disease.

*Value of* $T_2$

By incorporating information from the $T_2$-weighted image, the rate of false positive results may be decreased. Cysts, lymph nodes and certain types of myxomatous (benign tumour) are high in signal on the $T_2$-weighted images. Some malignant disease may also enhance on the $T_2$-weighted images. However, if the lesion is high in signal on the $T_2$-weighted imaging, it is more likely to be benign than malignant (Morris and Liberman, 2005).

*Kinetics*

Enhancement kinetics can be particularly helpful if the lesion has benign morphometric characteristics. In the case of a well-defined mass that could quite possibly be benign, enhancement kinetic data may help one decide whether a biopsy is required or whether it is safe to recommend a follow-up imaging of the lesion. The contrast enhancement of a lesion is related to the amount of contrast agent in the tissue. Pharmacokinetic models describe the contrast enhancement in the tissue using just a few free parameters. Such a model may give an indication of the behaviour of the contrast agent in the tissue and thus indicate if the tissue is benign or malignant (Furman-Haran and Degani, 2002).

*Automatic classification of suspicious lesions in the breast*

Recent research implies that CAE can, indeed, improve the interpretation of DCE-MRI of the breast. The techniques that are used to classify suspicious lesions in the breast are usually based on statistical and mathematical tools from the field of statistical pattern recognition and are aimed at objectively describing the difference between malignant and benign lesions, based on predefined features. These features are commonly derived from the morphological and kinetic features that are in the ACR-BIRADS lexicon (see Appendix C).

## 2.3 Statistical Pattern Recognition Overview

This section provides an overview of statistical pattern recognition. The primary goal of pattern recognition is the classification of data samples in a supervised or unsupervised manner. Among the different pattern recognition frameworks, statistical pattern recognition has been the one studied the most (Jain et al., 2000). The process of pattern classification consists of the following stages: definition of classes, pattern

representation, feature extraction and selection, classifier design and training, performance evaluation and possibly various pre- and post-processing stages (Jain et al., 2000). Classification problems are also challenging, because the performance of the final classification may only be as good as the stage with the lowest performance in the system. Thus, if the representation of the classes, the selected features or the design of the classifier is suboptimal, so will be the performance of the resulting classification system.

There are four principal approaches to pattern recognition (Jain et al., 2000):

1. Template matching. Using a template or a prototype of the pattern that is to be recognised. Each new pattern is then matched against the stored template using some similarity measure, while accounting for a degree of variability.
2. Statistical approach. In this approach, each pattern is represented as a point in a multi-dimensional feature space. The goal is to select a set of features that will provide sufficient separability (i.e. a simple discrimination boundary) between the different classes. Given a set of samples from each class (training set) decision boundaries are then constructed by a classifier in the selected feature space. These decision boundaries will determine the class of a new unseen pattern.
3. Syntactic approach. The syntactic approach is a hierarchical approach that is aimed at problems that include complex patterns. In this approach, each pattern is viewed as consisting of simpler sub-patterns in a hierarchical fashion. The lowest level patterns (i.e. the simplest sub-patterns) are called primitives and the complex patterns are then represented in terms of interrelationships between the primitives.
4. Neural networks. This is a massively-parallel computing system consisting of many simple processing units (perceptrons) with many interconnections. This method is capable of solving complex nonlinear discrimination problems owing to the flexibility that can be implicitly configured into the network.

Of these, the statistical pattern recognition is the most studied (Jain et al., 2000) and is often used in medical image analysis. Also, it provides a large variety of classification tools (i.e. classifier types) that can be selected to fit the type of problem and level of complication that is required for the solution (e.g. linear or non-linear).

### 2.3.1 **Statistical pattern recognition**

In statistical pattern recognition, a pattern is represented by a feature vector. The decision rules between the different classes are designed using concepts from a well-known statistical decision theory (Jain et al., 2000). When the class conditional densities are known, a Bayes decision rule (i.e. a decision rule that minimises a posterior expected value) can be applied for classifying the patterns. When the conditional density is unknown, however, a supervised or unsupervised classification method can be applied. In the supervised classification, the classification process consists of two main stages, training and testing (classification). In the training stage, a set of labelled patterns from each class is given to the classifier (i.e. the discrimination function) and the classifier is then adjusted to yield a different, but consistent, value for each class (e.g. 0 for class A and 1 for class B). It is then assumed that when a new pattern is presented to the classifier (from the testing set) the response value of the classifier will indicate to which class the new pattern belongs.

In the unsupervised classification, no training set is used. Instead, it is assumed that the structure of the classes will be automatically described by the classifier. In some cases, it is even assumed that the number of classes is unknown. In this research, however, only the supervised classification has been used, because both the number and description of the classes is known (benign vs. malignant).

### 2.3.2 **The 'curse of dimensionality'/'Peaking phenomenon'**

The quality of the classification strongly depends on the choice of features and type of classifier. A 'good' feature space is one that clearly separates classes. In other words, patterns from different classes in the feature space can be discriminated using a simple discrimination function (e.g. linear). The selected classifier needs to fit well to the training data (i.e. produce good separation between the classes) while, on the other hand, being as robust as possible to unseen data (generalisability). However, the performance of the classifier also depends on the relationship between the number of features and the size of the training set. In the naïve lookup table approach (i.e. partitioning the space into cells and associating each cell with a class), the size of the training set would be an exponential function of the feature space dimension (Jain et al., 2000). In less naïve approaches, the ratio may be smaller and less training samples are

required. Nevertheless, the number of training samples should always be much higher than the dimension of the feature space. This phenomenon is known as the 'curse of dimensionality'. A rule of thumb is that the number of training samples per class should be at least ten times the number of features (Jain et al., 2000). If the number of features is increased beyond this point, the performance of the classifier will usually degrade. This behaviour is known as the 'peaking phenomena'. The reason for this behaviour is that using a training set that is too small (relative to the number of features), yields an 'over-fitted' classifier that will not generalise to unforeseen data samples and thus have a low performance on a testing set.

The curse of dimensionality thus implies that the elimination of redundant information from the feature space (e.g. by reducing its dimensionality) will lead to a better classification quality (i.e. more robust), which motivated the development of dimensionality reduction techniques, such as the Principal Component Analysis (Webb, 2003, Moore, 1981, Jain et al., 2000) and the Independent Component Analysis (Webb, 2003, Jain et al., 2000, Comon, 1994). Some classifiers incorporate dimensionality reduction as an integral part, such as the Eigen-faces (Turk and Pentland, 1991) and the Anti-faces (Keren et al., 2000, Keren et al., 2001) classifiers.

### 2.3.3 **Classifiers**

Once a proper representation in a feature space, has been chosen for the patterns, a classifier should be trained and tested on the data. The role of the classifier is to implement a decision rule that will indicate to which class a given pattern belongs. In other words, it operates a discriminant function on the data. In the two-class case, the classifier creates a decision boundary in the feature space on which the discriminant function yields a certain value that is a pre-determined threshold. Patterns for which the discriminant function yields a value higher than this threshold are classified as class A, while the rest are classified as class B.

A brief overview of several types of classifiers that were used in this research follows. The review is based on (Webb, 2003).

### *Linear discrimination classifiers*

Let us assume a set of training patterns $x_1, \dots, x_n \epsilon \mathbb{R}^m$, where each pattern is assigned to one of two classes, $C_1$ or $C_2$. We now seek a vector $w \epsilon \mathbb{R}^m$ and a threshold $\omega_0 \epsilon \mathbb{R}$ such that

$$w^T x_i + \omega_0 > 0 \leftrightarrow x_i \in C_1$$

and

$$w^T x_i + \omega_0 < 0 \leftrightarrow x_i \in C_2$$

Ideally, we would like $w^T x + \omega_0$ to be positive for all the patterns of class $C_1$ and negative for the patterns of class $C_2$ (Webb, 2003). In practice, it is usually not the case and thus we seek to minimise the number of misclassifications. To do that, additional criteria should be employed. Some of the criteria used for linear discrimination are the Perceptron criterion, Fisher's criterion, Baysian criterion and the least mean squared error.

### *Logistic discrimination*

Logistic discrimination, or logistic regression, is based on the assumption that the difference between the logarithms of the density functions of the classes is a linear function of the variables *x*:

$$\log\left(\frac{p(x|C_1)}{p(x|C_2)}\right) = \beta_0 + b^T x$$

This assumption is satisfied by a variety of distributions and has been found to be useful in a wide range of real datasets that depart from normality (Webb, 2003). The discrimination rule of logistic discrimination is based on the ratio between the probabilities and is given by:

$$\frac{p(C_1|x)}{p(C_2|x)} \begin{cases} > 1 \rightarrow x \in C_1 \\ < 1 \rightarrow x \in C_2 \end{cases}$$

### *Support vector machines*

Support vector machines (SVMs), originally proposed by Vapnik (Cortes and Vapnik, 1995), construct a maximal margin separating hyper plane in the feature space to create a 'best' separation between the classes (Figure 2.8). The discriminant function that the

SVM constructs yields a corresponding label with value of +1 on one class and a label value of -1 on the other class. The SVM can be either linear or nonlinear. In the nonlinear case, the boundary between the classes will not be a plane.

In the linear case, the discriminant function of the SVM is given by:

$$g(x) = w^T x + \omega_0$$

with the decision rule:

$$w^T x_i + \omega_0 \begin{cases} > 0 \rightarrow x_i \in C_1 \; with \; corresponding \; value \; y_i = +1 \\ < 0 \rightarrow x_i \in C_2 \; with \; corresponding \; value \; y_i = -1 \end{cases}$$

In the nonlinear case, a nonlinear kernel function, $\phi(X)$, is embedded into the discriminant function. The discriminant function then becomes:

$$g(x) = w^T \phi(X) + \omega_0$$

which allows a more complicated decision boundary between the classes. Some of the commonly used kernels for SVM are Polynomial, Gaussian and Sigmoid.

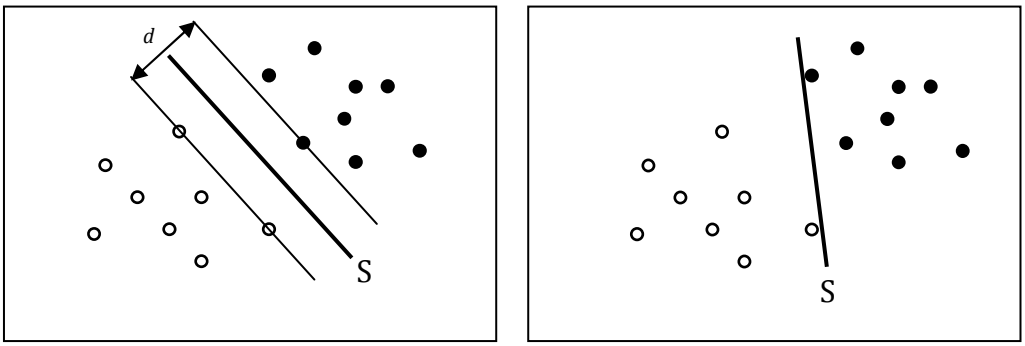



Figure 2.8: Two linearly separable sets of data with a separating hyperplane
The separating hyperplane on the left (marked with S) defines a maximum distance, *d*, between the margins (marked with thin lines). The patterns of each class that are on the margins are called 'support vectors'.

### 2.3.4 **Evaluating classification performance**

In real world classification systems, the classification error rate, $P_e$, is the ultimate measure of the performance of a classifier (Jain et al., 2000). The error rate of a classifier is estimated from the available sample patterns split into a training set and a test set. The classifier is first trained on the training set and then its performance is tested on the test set. The training set must be sufficiently large to avoid the curse of

dimensionality and to increase the robustness of the classifier, while the test set should be as large as possible to increase the confidence of the error rate estimation. It is also important that the training set and the test set be independent to avoid bias.

Given a set of classification results, let $t$ be the number of misclassifications out of $n$ samples in the test set. It can be shown that the probability density function of $t$ has a binomial distribution and the maximum likelihood estimates, $\widehat{P_e}$, of $P_e$ is given by $\widehat{P_e} = t/n$, with $E(\widehat{P_e}) = P_e$ and $var(\widehat{P_e}) = P_e(1 - P_e)/n$. Given that $\widehat{P_e}$ is a random variable, a confidence interval can be associated with it. This confidence interval shrinks as the number of samples, $n$, increases (Jain et al., 2000).

The Receiver Operating Characteristic (ROC) curve summarises the performance of a two-class discrimination rule, over a range of classification thresholds (Webb, 2003). The ROC curve is the plot of true positive (i.e. the proportion of patterns from class $C_1$ that were classified correctly) against the false positive (i.e. the proportion of patterns from class $C_2$ that were classified as $C_1$) by changing the decision rule threshold. In practice, the real ROC curve is unknown and is estimated by a trained classifier using a test set.

When measuring the performance of a classifier, often a single number is desired to represent its performance to simplify the comparison between the different classifiers. The area under the ROC curve (AUC) provides a single value, based on the ROC curve that indicates the potential performance of the classifier (Webb, 2003). A value of 1 means, that given the right threshold, the classifier will perfectly discriminate between the two classes. A value close to 0.5, on the other hand, means that for any threshold, the classifier will not be able to discriminate between patterns of the different classes with a high probability (i.e. larger than 0.5). The AUC measure has a few more important features. One is that it does not make a misclassification-cost assumption. When measuring the error rate $P_e$, we implicitly assume that the cost of misclassification is equal for false positive and false negative. In many cases, we do not have enough information to associate different costs with the different types of misclassification. The AUC measure, however, is free from this assumption. Another important feature of the AUC is that it is independent of prior knowledge about the distribution of samples. Thus, the AUC will be the same whether we train the classifier under the assumption the

population distribution between the different classes is 50% in each class or something else.

## 2.4 Summary and conclusions

This chapter has introduced the basic principles behind MRI in general and DCE-MRI specifically. It has also described the basic physiology of breast cancer and the anatomy of the human breast. The last section of the chapter has given an overview of statistical pattern recognition, including basic classifier types and methods for evaluating classification performance that will be used later, in Chapter 6. The next chapter of the thesis will discuss models of contrast enhancement for DCE-MRI, which are derived from the behaviour of the contrast agent in the human body after it is injected into the patient's bloodstream. In particular, a new model of contrast enhancement will be described and compared with the existing pharmacokinetic models of DCE-MRI.

MRI currently shows the most promise for the detection of breast cancer and provides a good compromise between 3D volume coverage, image acquisition time, spatial resolution and temporal sampling (Heiberg et al., 1996). However, although sensitivity is high, specificity of MRI is still low (Morris and Liberman, 2005, Warren and Coulthard, 2002) and the resulting images often suffer from low SNR and bias field effects (Hayton, 1998).

# 3. Parametric models of contrast enhancement

As discussed in Chapters 1 and 2, the amount of data that needs to be interpreted in DCE-MRI is often huge, and is likely to increase in the future. Thus, radiologists can be overwhelmed by the amount of data and the increasing work load. Parametric models of enhancement provide ways to visualise the data more efficiently, decrease the size of the data and permit the development of kinetic features for the automated classification of breast lesions (the subject of Chapter 6).

Existing approaches vary between using the raw data, its derived statistics and a range or empiric kinetic and pharmacokinetic models. This chapter introduces the reader to the motivation and types of parametric models in DCE-MRI. It reviews existing models of enhancement and presents a novel parametric model of enhancement for DCE-MRI of the breast. Also, it presents an empirical evaluation of the goodness-of-fit of four parametric models of contrast enhancement for DCE-MRI of the breast: the Tofts (Tofts and Kermode, 1991), Brix (Brix et al., 1991), and Hayton (Hayton et al., 1997) pharmacokinetic models, and the proposed empiric model. Each of these models has three free parameters. The aim of this work was to analyse the behaviour of each model under different optimization conditions to determine which one had the best performance for each given optimisation system. The core material of this chapter was published in (Gal et al., 2007b).

## 3.1 Introduction

Parametric models of contrast enhancement reduce the dimensionality of DCE-MRI by transforming a dataset of typically 5-9 time points (i.e. volumes) to two or three parameters. This conversion helps standardise the size of the data (to have a constant size, instead of a varying number of volumes) and makes the data easier to visualise (by

using the parameter values from the fitted model) (Mehnert et al., 2005, Vidholm et al., 2007). Also, computer-assisted evaluation often derives information from such models to segment or classify suspicious lesions.

In most protocols used with DCE breast MRI, the injection is applied as a bolus (i.e. rapidly injected intravenous injection). Variations in injection time, in addition to the lag times between the injection and the imaging need to be considered when analysing the shape of the enhancement curves. Contrast enhancement patterns help to differentiate the benign from the malignant lesions in DCE-MRI of the breast (Furman-Haran and Degani, 2002). These are particularly helpful when the morphometric features of a lesion make the differentiation, and thus the indication if the tissue is benign or malignant, difficult. The behaviour of the contrast agent in the body tissues can be described using a model-based equation that approximates the entry and washout of the agents in the body.

Gd-based contrast agents cannot cross the cell membranes and enter the cells. Therefore, many models assume that the contrast agent is distributed between two main tissue compartments, the intra-vascular plasma volume and the extra-vascular/extra-cellular volume (Figure 2.4). Many models also share several additional assumptions, usually related to the water exchange between the tissue compartments and water interaction with the contrast agent. Those assumptions usually lead to extra parameters that are experimentally tested or are estimated by the model fitting process.

Several pharmacokinetic models for breast DCE-MRI have been developed to date including, Tofts (Tofts and Berkowitz, 1994, Tofts and Kermode, 1991), Brix (Brix et al., 1991), Larsson (Larsson et al., 1990), and Hayton (Hayton et al., 1997). Pharmacokinetic models make assumptions about the contrast agent's perfusion process and the water exchange rates between the prescribed tissue compartments. As a result of these assumptions, some prior knowledge is typically required to fit the model to the measured DCE-MRI data. This prior knowledge is usually introduced into the model as additional parameters that have to be estimated for each study or statistically measured from previous studies (Furman-Haran and Degani, 2002, Tofts and Kermode, 1991, Larsson et al., 1990, Tofts et al., 1995). This can introduce additional noise and/or bias in the model-fitting process. In theory, one of the advantages of the pharmacokinetic models over the empiric models is that they make

possible inter-study comparability. However, in practice, this is rarely the case, because of factors such as, differing spatial and/or temporal resolutions, $T_1$ contrast dynamic range, the ability to quantify $T_1$, and the ability to measure the arterial input function (Jackson et al., 2005).

## 3.2 Review of parametric models of enhancement

Existing parametric models of contrast enhancement can be broadly classified into two groups, model-based parametric models and empiric parametric models (Furman-Haran and Degani, 2002). The former are pharmacokinetic models that mathematically characterise the change in the concentration of the contrast agent over time in terms of physiologic properties, whilst the latter are objective parameterisations that describe the enhancement kinetics without specific physiologic interpretation. An overview of parametric models of contrast enhancement can be found in (Furman-Haran and Degani, 2002). This section reviews a variety of parametric models of contrast enhancement and methods of fitting these models to DCE-MRI data. The review focuses on models with only three free parameters, because models with more than three free parameters, such as the Larsson model (Larsson et al., 1990), require data with a high temporal resolution.

### 3.2.1 Pharmacokinetic models

This section reviews models that are based on compartmental contrast-agent diffusion in the body. Data providing the time-dependent changes in the concentration of the contrast agent in body tissues can be analysed by nonlinear regression methods yielding parameters that have characteristic tissue properties (Hittmair et al., 1994, Morris and Liberman, 2005, Tofts and Berkowitz, 1994). Various models yield different physiologic parameters that may cause different fitting of the data to different models.

The pharmacokinetic models developed for DCE-MRI of the breast are each formulated as a sum of exponential functions (arising from the solution of the differential equations governing the exchange rate of the contrast agent between the prescribed tissue compartments) describing the concentration of a contrast agent in the tissue as a function of time. Each model seeks to describe the relationship between the intensity in the image, as a function of time, and the concentration of the contrast agent in the observed tissue. There is empirical evidence suggesting that the relationship between

the $T_1$ MR relaxation time and the concentration is approximately linear (Hittmair et al., 1994). This, then, implies that the concentration at time *t* is directly proportional to the *relative signal increase* at time *t* (Hayton, 1998, Chen et al., 2005). Given that the assumption is only approximately true, the accuracy of these models is limited and their closed forms are quite rigid and will always be a sum of exponentials.

The Tofts model was the first pharmacokinetic model proposed for breast MRI. It is an adaptation of a pharmacokinetic model initially proposed by Tofts and Kermode (Tofts and Kermode, 1991) for the measurement of the blood-brain barrier permeability. This model (which measures signal enhancement) can be described as follows (Furman-Haran and Degani, 2002, Tofts and Kermode, 1991):

$$S_E(t) = \frac{(1-e^{-\frac{T_R}{T_{10}}}\cos\alpha)(1-e^{-T_R(\frac{1}{T_{10}}+R_1C_t(t))})}{(1-e^{-T_R(\frac{1}{T_{10}}+R_1C_t(t))}\cos\alpha)(1-e^{-\frac{T_R}{T_{10}}})}, \tag{3.1}$$

where $C_t(t)$, the concentration of the contrast agent, is given by:

$$\boldsymbol{C_t(t) = D\left(b_1 e^{-m_1 t} + b_2 e^{-m_2 t} + b_3 e^{-\frac{Kt}{v_1}}\right)}, \tag{3.2}$$

where:

$b_1 = \frac{K a_1}{\frac{K}{v_1} - m_1}$

$b_2 = \frac{K a_2}{\frac{K}{v_1} - m_2}$

$b_3 = -(b_1 + b_2)$

D is the contrast agent dose (m *M* / kg body wt)

$a_1, m_1$ and $a_2, m_2$ are the corresponding amplitude and decay constants of the contrast agent in the body tissue and blood plasma (vary between contrast agents and between different species — these constants are measured experimentally)

$v_1$ is the extra-cellular volume fraction, available to the contrast agent (practically unknown)

$K$ is the permeability factor (practically unknown)

$R_1$, $R_2$ are the relaxivities of the contrast agent (of $T_1$, $T_2$ respectively)

$T_{10}$ is the intrinsic tissue relaxation of $T_1$ (i.e. the relaxation time, before injection, which is practically unknown)

$T_R, T_E, \alpha$ are parameters of the acquisition (Time of pulse repetition, Time of pulse echo, Initial pulse angle, respectively).

From the definition of the Tofts model it can be seen that to fit the model to a set of points, a non-linear fitting method should be used for the three parameters ($K$, $v_1$, $T_{10}$), whilst the values for the other parameters are evaluated experimentally ($a_1, m_1$ and $a_2, m_2$) and may vary from one subject to another.

In certain cases, the lack or inaccuracy of the information about the value of certain parameters may result in an ill posed problem that leads to a poor fitting rate and a high sensitivity to noise.

Several additional models have since been proposed including the Brix model (Brix et al., 1991) and the Hayton model (Hayton et al., 1997). These two are of high interest, because they have no more than three free parameters. The Brix model can be expressed by the following explicit expression for a spin echo sequence:

$$\boldsymbol{S(t) = S(0) + S(0) \cdot A \cdot \left[v\left(e^{k_{el}t'} - 1\right)e^{-k_{el}t} - u\left(e^{k_{12}t'} - 1\right)e^{-k_{21}t}\right]}, \quad (3.3)$$

where:

$A$ is a parameter that depends on the properties of the tissue, the MR sequence and the infusion rate (practically unknown)

$v = [k_{el}(k_{21} - k_{el})]^{-1}$

$u = [k_{21}(k_{21} - k_{el})]^{-1}$

$k_{el}$ is a pre-measured systematic constant

$k_{21}$ and $k_{12}$ are free parameters

$t'$ is a time parameterization that corrects for the offset between the injection of the agent and the time it enters the lesion. During the contrast agent infusion ($0 \leq t \leq \tau$) the identity $t' = t$ has to be used; afterwards, the identity $t' = \tau$ has to be used.

The Hayton model (Hayton, 1998, Hayton et al., 1997) has a more parsimonious form. It contains three free parameters and can be expressed as a difference between two exponentials:

$$\boldsymbol{C(t)} = \frac{\boldsymbol{A}}{\boldsymbol{a-b}}\left(\boldsymbol{e^{-bt}} - \boldsymbol{e^{-at}}\right), \tag{3.4}$$

where $A$, $a$ and $b$ are free parameters and are constrained to be positive.

Although most of these models have been shown to be theoretically compatible, historically there exist significant differences in the assumptions made in the application of these models (Jackson et al., 2005). One example is the manner in which the arterial input function (AIF[1]) is determined. In most study protocols for breast DCE-MRI, the contrast agent is administered as a bolus either manually or using a power injector. Therefore, the AIF and the lag time between injection and imaging need to be considered when analysing the shape of the enhancement curves. The early Tofts and Kermode model assumes an AIF of bi-exponential form. The Brix model assumes a patient-specific exponentially decaying AIF whose parameter is included as an additional free parameter in the model. The Hayton model assumes a bi-exponential model (if the bolus injection is assumed to be instantaneous then the Hayton model has three free parameters).

However, it is important to realise that physiological parameters derived from a fit of one model are unlikely to be comparable to those derived from a fit of another model to the very same data (Jackson et al., 2005). The pharmacokinetic models developed for DCE-MRI of the breast are each formulated as a sum of exponential functions describing the concentration of a contrast agent in the tissue as a function of time.

### 3.2.2 **Empiric models**

In addition to the development of the pharmacokinetic models, other models of enhancement were developed. Those models do not use prior knowledge about the behaviour of the contrast agent in the body tissue but rather, try to present a model that

[1] AIF describes the concentration of contrast agent in the arterial blood, feeding a tissue, as a function of time.

will fit best to the observed data with a minimum number of free parameters. These include the Fan model (Fan et al., 2004), which consists of a product of three exponential terms and five free parameters. Another model of enhancement is the three-time-point model, proposed by Degani *et al.* (Degani et al., 1997). This model is based on selecting three time points (where one is always the pre-contrast) to represent the data. Other empiric methods include the 'Method of Factor Analysis' and 'Triple Acquisition Rapid Gradient Echo Technique' (Furman-Haran and Degani, 2002).

### 3.2.3 **Model fitting methods**

Once a model of contrast enhancement is given, it is required to fit the model to the data to evaluate its parameter values. This can be done on a pixel-by-pixel basis (i.e. for each enhancement curve in the data) or on a sum of enhancement curves, taken from a region of interest. To fit a model to a sampled DCE curve, a non-linear fitting method is needed. Both deterministic and stochastic algorithms can be used. The objective function used for the fitting process is usually the least squares. A few common deterministic methods of fitting can be found in the literature, such as Levenberg-Marquardt, Steepest Descent, Newton-Gauss, Nelder-Mead Simplex and Trust Region (Buckley et al., 1994, Sykulski et al., 1998). Stochastic algorithms include Monte-Carlo, Genetic Algorithms and Simulated Annealing (Sykulski et al., 1998).

Comparison between non-linear fitting methods is strongly influenced by the type and structure of the problem, thus it is very limited and mainly restricted to '*hand picked*' types of problems. Sykulski *et al.* suggest that although deterministic methods are generally robust and computationally economical, they often fail to reach the global maximum/minimum (Sykulski et al., 1998). Of these, the Levenberg-Marquardt has been found the most efficient and is commonly used for fitting parametric models of contrast enhancement in DCE MRI (Ahearn et al., 2005, Hayton, 1998). Of the stochastic methods, Genetic Algorithms have proven useful for electromagnetic applications. The Simplex optimization method, however, has been found to be less sensitive to initial conditions than the Levenberg-Marquardt algorithm, although this conclusion was drawn from a comparison that was made using noisy synthetic data generated from a simple kinetic model with Gaussian noise (Buckley et al., 1994).

Non-linear fitting methods are usually time-consuming, especially when a high level of precision in the final solution is required (i.e. closeness to global minimum). Anne L. Martel (Martel, 2006) recently proposed a way of significantly accelerating the fit of pharmacokinetic models to DCE curves by creating a pre-calculated map of the model parameter values. The intensity curves from the DCE-MRI data are compared to those in the map and the parameter values of the most similar curve in the map are selected. Principal component analysis (PCA) is used to reduce the dimensionality of the problem and to further accelerate the calculations.

Given that a DCE-MRI data set typically comprises only a small number of time points, e.g. 5–9, spaced a minute or more apart, and that the data can be extremely noisy, it is dangerous to attempt to fit a multi-exponential model with a similar number of parameters. It can be shown that the error in the fit, $E$, is given by (Collins and Padhani, 2004):

$$\boldsymbol{E} = \sqrt{\sum_{i=1}^{N} \frac{\left(C_i^2 - c_i^2\right)^2}{N-P}}, \tag{3.5}$$

where $N$ is the number of time points, $P$ is the number of free parameters in the model, $C_i$ is the concentration of contrast agent, and $c_i$ is the model estimate of the concentration of contrast agent. Consequently, for a fixed number of parameters, the fewer the number of data points the larger the error. In practice it is not possible to increase the number of time points without sacrificing spatial resolution or reducing the field of view to only part of the breast volume. Neither option is acceptable. Instead, it is preferable to diminish the error by reducing the number of parameters. One possibility is to attempt to measure a parameter experimentally rather than to try and estimate it. However, this introduces measurement errors into the model. An alternative is to build an empiric model that will imitate the contrast enhancement in DCE-MRI without relying on physiological characteristics and with a small number (i.e. up to three) of free parameters.

## 3.3 New empirical model of enhancement

Our proposed model is not derived from pharmacokinetics but rather, chosen empirically to seek a parsimonious fit to the time series data, while still having the flexibility to model a variety of enhancement curve shapes typically observed in DCE-

MRI (Figure 1.1). The model requires a fit of three free parameters. It is free of experimental errors and the pre-measurement of physiological properties because ,(i) it does not rely on measured parameters or constants relating to the type or density of the tissue and, (ii) it does not assume a particular relationship between the observed change in signal intensity and the concentration of contrast agent. The model is defined in terms of only one exponential function and the range of its parameters can be restricted to a relatively narrow set of values. The proposed model of the voxel-wise tissue signal intensity over time is given by:

$$\boldsymbol{f(t) = a \cdot t \cdot e^{-\frac{t^c}{b}}}, \tag{3.6}$$

where $t$ is the time that has elapsed since the injection of the contrast agent; and $a$, $b$ and $c$ are positive, real parameters. When $c \approx 0$ the model approximates a straight line. Given that $t = 0$ is the acquisition starting point, we do not deal with values of the model when $t < 0$.

### 3.3.1 **Analytical properties of the model**

To find the peak of the curve (i.e. point of maximum intensity), the first derivative of the model has to be computed:

$$\boldsymbol{f'(t) = a \cdot e^{-\frac{t^c}{b}} \left(1 - \frac{t^c \cdot c}{b}\right)} \tag{3.7}$$

The model has a single maximum at the (wash-in/wash-out) *critical point* (i.e. peak) for which $f_t'(t) = 0$:

$$\boldsymbol{t_{peak} \coloneqq arg\,\max_t f(t) = \left(\frac{b}{c}\right)^{\frac{1}{c}}} \tag{3.8}$$

And the value at the peak is thus:

$$\boldsymbol{f\left(t_{peak}\right) = a\left(\frac{b}{c \cdot e}\right)^{\frac{1}{c}}} \tag{3.9}$$

### 3.3.2 **Motivation for the form of the model**

The model is a product of three factors:

1. The first factor is the coefficient $a$, which is the scaling parameter. Given a value, $V(t)$, of the form

$$V(t) = t \cdot e^{-\frac{t^c}{b}},$$

the value of enhancement at the peak,

$$V(t_{peak}) = \left(\frac{b}{c \cdot e}\right)^{\frac{1}{c}}$$

can be controlled by the coefficient $a$ alone. This allows the model to be fitted to any magnitude of enhancement, regardless of the curve's shape.

2. The second factor is $t$. This factor makes sure that for $t = 0$ the value of the model will be 0:

$$f(0) = a \cdot 0 \cdot e^{-\frac{0^c}{b}} = a \cdot 0 \cdot 1 = 0$$

Once $t > 0$ the value of the model will increase as long as $t < t_{peak}$.

3. The third factor is the exponential term $e^{-\frac{t^c}{b}}$. It is responsible for the shape of the curve. The parameter $b$ determines the width of the Gaussian function and thus it influences both the wash-in and the wash-out slopes. The parameter $c$ is a bias factor that influences the difference between the two slopes.

### 3.3.3 **Fitting the model to clinical data**

To fit the model to the enhancement curve of a given voxel, a non-linear fitting algorithm is required. The minimization problem, in the least squares sense, is not convex and thus a global best fit is not guaranteed. Suitable algorithms include the Levenberg-Marquardt (LM) algorithm (Marquardt, 1963) and the Nelder-Mead (NM) simplex algorithm (Nelder and Mead, 1965) (both non-linear least squares optimisation algorithms). Our experience is that the fitted values for the model parameters lie within a relatively narrow range, e.g. based on several clinical data sets additional to those described in Experiment 1 (Table 3.1), we observed that, $a \in (0,200)$, $b \in (0,100)$, $c \in (0,3)$.

## 3.4 Empirical evaluation of the proposed model

To evaluate the ability of the proposed model to describe DCE-MRI data, its goodness-of-fit (GOF) was compared against three other models. The comparison included the following models of contrast enhancement for DCE-MRI: the Tofts, Brix, and Hayton pharmacokinetic models, and a novel empiric model. The goodness-of-fit of each model was evaluated on breast MRI data from a local radiology practice in two experiments: (i) comparison of two model-fitting algorithms (Levenberg-Marquardt and Nelder-Mead) and two fitting tolerances and, (ii) goodness-of-fit with respect to the temporal resolution.

### 3.4.1 Experiment 1: Evaluation of the goodness-of-fit with respect to the fitting algorithm and the fitting tolerance

The aim of this experiment was to evaluate the impact that the choice of the non-linear least squares fitting algorithm — the LM algorithm and the NM simplex algorithm — and the choice of fitting tolerance level had on the ability of the Tofts, Brix, and Hayton pharmacokinetic models and the proposed empiric enhancement model to fit clinical DCE-MRI data. Three DCE-MRI data sets, originating from routine clinical breast MRI examinations, were used for this experiment, one with malignant enhancement, one with benign enhancement, and one with no suspicious enhancement (findings of the reporting radiologist). The data were acquired on a 1.5T Signa EchoSpeed (GE Medical Systems, Milwaukee, USA). The scan/sequence details are given in Table 3.1. The GOF of the models was measured using two metrics, (i) the coefficient of determination R-squared and, (ii) the mean squared error (MSE). The R-squared was chosen because of its insensitivity to scale, thus yielding an assessment of the GOF that is not influenced by the scale of enhancement in the data, and thus allows comparing different tissue types. The MSE was chosen because it measures directly the goal of the optimization algorithm and assesses how well the algorithm achieved this goal. Thus, it allows the comparison of different optimisation algorithms, when using the same data set.

Model-fitting was performed voxel–wise. Starting values for the model parameters were defined to be the average values of these parameters estimated in model-fitting experiments on several additional clinical data sets. The voxel-wise GOFs were averaged over, (i) the entire breast volume and, (ii) a selected region of interest (either the area of

enhancement identified by the reporting radiologist or an arbitrary area of enhancement in the case with no suspicious enhancement). The resulting mean GOFs were then averaged over the three data sets. Two tolerance levels were selected to represent, (i) 'quick and dirty' fitting ($10^{-3}$) and, (ii) accurate fitting ($10^{-6}$). It should be noted, however, that the meaning of the tolerance levels is different for the two algorithms; for the LM algorithm, the tolerance level is defined with respect to the optimization function, and for the NM algorithm, it is defined with respect to the model parameters. The results are presented in Table 3.2 and Table 3.3.

Table 3.1: Scan/sequence details for the data sets used in Experiment 1

| | |
|---|---|
| Acquisition matrix | 256×256 |
| Image matrix | 512×512 (†ZIP512) |
| Flip angle | 10° |
| No. sagittal slices | 40 |
| Slice thickness | 2.5 mm (†ZIP2) |
| No. post-contrast volumes | 5–9 |
| Pixel spacing | 0.39 mm |

†ZIP512 and ZIP2 are image resizing algorithms that use interpolation

### 3.4.2 **Experiment 2: Evaluation of the goodness-of-fit with respect to missing temporal data**

Experiment 1 evaluated how well each model fits to real data. However, although the proposed model fits better to the data than the other models, one may still relate this advantage to the flexibility of the model. For example, by taking a polynomial function of order $N$-1 or higher (where $N$ is the number of time points in the data), we can create a model that will perfectly fit any given data. To check how well the model represents reality, it is necessary to check how well it can predict data points that are missing from the time series, but can be inferred using physiological prior knowledge (or measured, using a higher temporal resolution). Hence, the aim of this experiment was to evaluate the impact that temporal resolution has on the ability of the models considered in Experiment 1 to predict the behaviour of real DCE-MRI data. A custom high temporal resolution DCE-MRI data set was acquired for this experiment: 26 volumes (15 seconds per volume), 8 slices (5 mm thickness) per volume, and 256×256 pixels per slice (0.78 mm square pixels). The data were obtained from a patient with a previously-identified malignant lesion. A high temporal resolution was achieved without compromising the spatial resolution by restricting the field of view to the vicinity of the lesion.

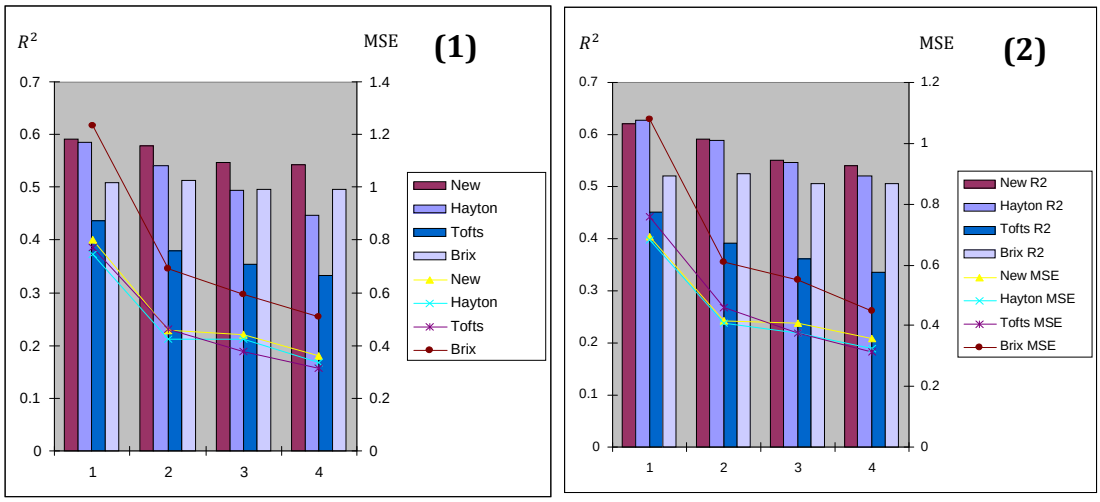
Model-fitting was performed voxel-wise using the LM and NM algorithms and the two tolerance levels defined in Experiment 1. The GOFs were measured for the complete time series as described in Experiment 1. As in Experiment 1, the GOFs were averaged over, (i) the entire breast volume and, (ii) a region of interest (ROI) within the malignant enhancement. Next, the data were temporally sub-sampled, the model-fitting repeated, and the GOFs computed with respect to the complete time series. The method used for sub-sampling involves taking the first volume out of every $N$ volumes and discarding the rest. Sub-sampling levels of $N$ = 2, 3, and 4 were used ($N$ = 1 is the complete time series). The results are presented in Figure 3.1.

Table 3.2: Experiment 1: Average GOF for the Levenberg-Marquardt algorithm (best results are highlighted in bold)

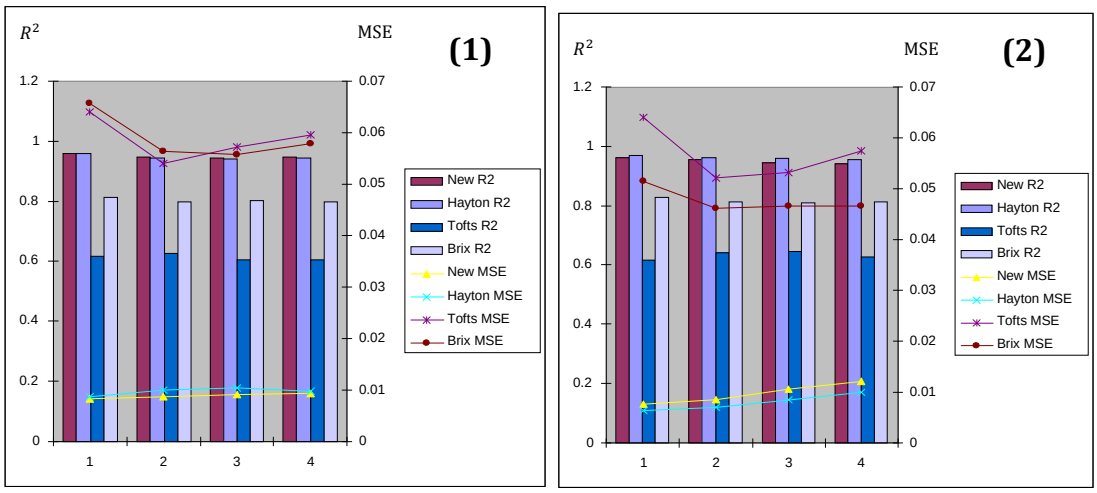
| Model name | Tol. Level | Whole Breast | | Enhancing Region | |
|---|---|---|---|---|---|
| | | $R^2$ | MSE | $R^2$ | MSE |
| New | $10^{-3}$ | **0.797** | 0.728 | **0.955** | **0.010** |
| | $10^{-6}$ | **0.826** | 0.657 | **0.967** | **0.007** |
| Hayton | $10^{-3}$ | 0.773 | **0.699** | 0.953 | **0.010** |
| | $10^{-6}$ | 0.817 | **0.654** | 0.964 | 0.008 |
| Tofts | $10^{-3}$ | 0.651 | 5.603 | 0.821 | 0.080 |
| | $10^{-6}$ | 0.652 | 2.125 | 0.821 | 0.042 |
| Brix | $10^{-3}$ | 0.716 | 1.059 | 0.890 | 0.029 |
| | $10^{-6}$ | 0.729 | 0.950 | 0.898 | 0.026 |

Table 3.3: Experiment 1: Average GOF for the Nelder-Mead algorithm (best results are highlighted in bold)

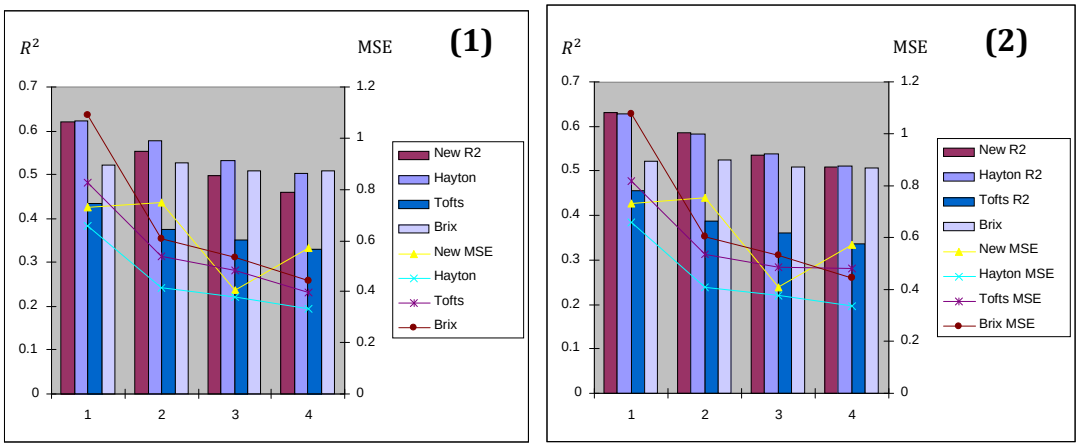
| Model name | Tol. Level | Whole Breast | | Enhancing Region | |
|---|---|---|---|---|---|
| | | $R^2$ | MSE | $R^2$ | MSE |
| New | $10^{-3}$ | 0.8241 | 1.1990 | **0.9680** | **0.0068** |
| | $10^{-6}$ | **0.8385** | 1.1315 | **0.9680** | **0.0068** |
| Hayton | $10^{-3}$ | **0.8253** | **0.9736** | 0.9647 | 0.0078 |
| | $10^{-6}$ | 0.8273 | **0.9727** | 0.9649 | 0.0077 |
| Tofts | $10^{-3}$ | 0.6504 | 1.2364 | 0.8211 | 0.0419 |
| | $10^{-6}$ | 0.6504 | 1.2364 | 0.8211 | 0.0419 |
| Brix | $10^{-3}$ | 0.7326 | 1.0036 | 0.8976 | 0.0265 |
| | $10^{-6}$ | 0.7332 | 0.9996 | 0.8983 | 0.0261 |



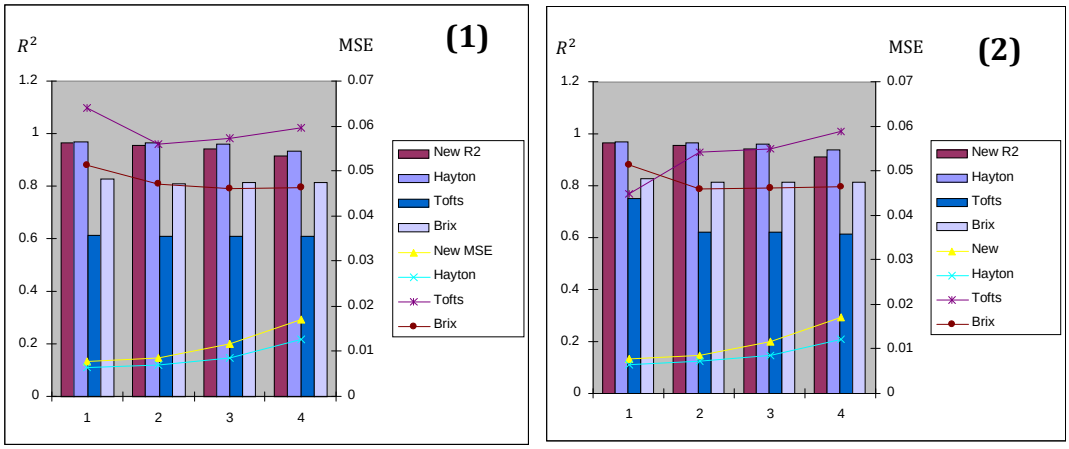

A. Whole breast – LM optimization




B. Region of interest – LM optimization




C. Whole breast – NM optimization




D. Region of interest – NM optimization

Figure 3.1: Experiment 2: The mean GOFs for the four enhancement models for sub-sampling levels N = 1, 2, 3, and 4

The horizontal axis in each graph represents the sub-sampling level, the bars show the mean R-squared values (the scale is on the left-hand side vertical axis), and the lines show the mean MSE values (the scale is on the right-hand side vertical axis). The left-hand column of graphs (1) correspond to a tolerance level of $10^{-3}$ and the right-hand ones (2) to a tolerance of $10^{-6}$.

### 3.4.3 **Discussion**

This section presented an empirical evaluation of the goodness-of-fit of four parametric models of contrast enhancement for DCE-MRI of the breast, the Tofts, Brix, and Hayton pharmacokinetic models, and a novel empirical model. The GOF of each model was evaluated with respect to, (i) two model-fitting algorithms (Levenberg-Marquardt and Nelder-Mead) and different fitting tolerances (Experiment 1) and, (ii) temporal resolution (Experiment 2). The results of Experiment 1 show that, irrespective of the fitting algorithm or whether the whole breast or a ROI is considered, a smaller termination tolerance leads to an improved GOF, albeit minor in the case of the NM algorithm. In addition, although the NM algorithm yields a slightly better R-squared values than the LM algorithm, it offers no consistent advantage over the LM algorithm in terms of the MSE values. Another observation is that the GOFs for the chosen ROIs (enhancing regions) are significantly better than those for the whole breast. This is likely because of the improved average signal-to-noise ratio within the ROI as compared to the breast as a whole.

The results of Experiment 2 for the whole breast show that, irrespective of the fitting algorithm and tolerance level, reduced temporal resolution leads to reduced mean R-squared values. In addition, they show that both the new model and the Hayton model yield better R-squared values than the other two models for both the whole breast and ROI, and better MSE values for the ROI. This implies that although Hayton's and the new model are more parsimonious than the other models, they tend to predict the behaviour of the contrast agent in the body more accurately.

Surprisingly, in the case of the whole breast, the MSE values generally decrease (improve) with decreasing temporal resolution. A possible explanation for this could be that, for the same stopping criterion, the algorithms may converge faster when there are fewer sample points (because the fitting problem is generally ill-posed).

### 3.4.4 **Conclusions**

The results of the experiments demonstrate that under the various optimization conditions considered, in general, both the proposed empiric model and the Hayton model fit the data about equally well and that both of these models fit the data better than the Tofts and Brix models. The advantages of the new model over the Hayton

model are that it has a more parsimonious form and that it makes no physiological assumptions (and thus its parameters cannot be linked to physiological properties). The simpler form of the new model and that, in practice, its parameters lie within a narrow range of values, suggests that it should be possible to fit this model faster than the Hayton model.

## 3.5 **Summary**

This chapter has reviewed models of contrast enhancements and methods for model fitting. The chapter has also presented a novel model of contrast enhancement and a comparison between the new model and three pharmacokinetic models from the literature, all with three free parameters. The new models are used in Chapter 4 for an evaluation of a new denoising algorithm, in Chapter 5 for the segmentation of suspicious lesions in the breast and in Chapter 6 for generating novel features for classification of suspicious lesions in DCE-MRI of the breast.

# 4. Denoising of DCE-MRI

This chapter deals with the issue of denoising DCE-MR images. Noise in MRI may mask fine details in the image and diminish its effective resolution, making both manual and automatic analysis of the image more complicated and less reliable. This chapter introduces the motivation for denoising methods for MRI. Also, the chapter presents a novel denoising method. The method is called Dynamic Non-Local Means (DNLM) and is designed specifically for the denoising of DCE-MRI. The DNLM is a variation of the Non-Local Means algorithm (NLM) (Buades et al., 2004, Buades et al., 2005b, Buades et al., 2005a, Buades et al., 2008). It exploits the redundancy of information in the temporal sequence of images whilst at the same time accounting for contrast enhancement. The chapter also presents quantitative and qualitative evaluations of the performance of the DNLM relative to seven other denoising methods (the Matlab code for the NLM and DNLM algorithms is given in Appendix C). The core material of this chapter was published in (Gal et al., 2009b, Gal et al., 2008, Gal et al., 2010).

## 4.1 Introduction

The interpretation of DCE-MRI data is a complex task for the radiologist, because of the large quantity of data that is acquired (several image volumes in a single scan). The task is further complicated by the presence of hardware-induced noise, geometric distortions and intensity non-uniformities (bias field), in addition to motion artefacts resulting from patient movement during image acquisition. Also, the noise in MRI is higher for shorter acquisition time. This may become a significant factor when acquisition time is strictly limited, such as in DCE-MRI. Moreover, some MRI procedures, such as fat suppression, reduce the SNR in the image. Ideally, the image data should have the highest possible signal-to-noise ratio (SNR) and the least artefacts for interpretation. This chapter focuses on the critical issue of noise removal in DCE-MRI. Sources of this noise include fluctuations in the receiver coil electronics and thermal noise from electrically-conducting tissue within the patient. A good denoising algorithm

will attenuate noise while, at the same time, preserving kinetic enhancement information and fine structural details, and minimising the introduction of new artefacts (e.g. smoothing artefacts). The DNLM algorithm exploits the redundancy of information in the temporal sequence of the images while, at the same time, reducing the propensity of the 'non-local' property of the NLM to create geometric artefacts. The DNLM algorithm is compared to seven other denoising methods: simple Gaussian filtering (Gonzalez and Woods, 2002), the original NLM algorithm (Buades et al., 2004), a trivial extension of the NLM to include the temporal dimension, bilateral filtering (Tomasi and Manduchi, 1998), anisotropic diffusion filtering (Perona and Malik, 1990), wavelet adaptive multiscale products threshold (Bao and Zhang, 2003), and traditional wavelet thresholding (Bao and Zhang, 2003, Pan et al., 1999).

The evaluations include quantitative evaluations on simulated and real data (20 DCE-MRI data sets from routine clinical breast MRI examinations) in addition to a qualitative evaluation by 24 observers (including 14 image/signal-processing experts and 10 clinical breast MRI radiographers).

## 4.2 **Modelling noise in DCE-MRI**

The noise in MR images has been studied and modelled, and has been found to be of Rician Distribution. In contrast to the Gaussian additive noise, the Rician noise is signal-dependent and is therefore more difficult to separate from the signal (Nowak, 1999). Macovski (Macovski, 1996) describes the different sources of noise in the MRI process as follows:

$$SNR = C \cdot f(Ob) \cdot g(Im) \tag{4.1}$$

where the first factor, C, represents the physical constants of the system that are beyond our control (e.g. magnetic flux, current in conductors); the second factor, $f(Ob)$, is a function of the dimensions of the object being imaged; and the third factor, $g(Im)$, is a function of the chosen imaging parameters. The last factor can be expressed as follows (Macovski, 1996):

$$g(Im) = \omega_0 V_h \sqrt{T}, \tag{4.2}$$

where $\omega_0$ is the readout frequency, $V_h$ is the voxel size, and $T$ is the imaging time. It is immediately apparent that as the voxel size is decreased (i.e. as resolution is increased) the SNR decreases proportionally. An effective denoising method is, therefore, of great interest, because it offers the clinician the possibility of either increasing the image resolution without increasing the effect of noise or having a better SNR for a given image resolution.

A statistical model of the noise in MR images was developed by Gudbjartsson and Patz (Gudbjartsson and Patz, 1996). According to their model, the distribution for a measured intensity, *M*, of a pixel can be described by:

$$P_M(M) = \frac{M}{\sigma^2} \cdot e^{-\frac{M^2+A^2}{2\sigma^2}} \cdot I_0\left(\frac{A \cdot M}{\sigma^2}\right), \tag{4.3}$$

where, *A* is the image pixel intensity in the absence of noise

$\sigma$ is the standard deviation of the Gaussian noise in the real and imaginary images (which assumed to be equal)

*M* is the measured intensity

$I_0$ is the modified zero order Bessel function of the first kind.

Gudbjartsson and Patz point out that for a small SNR ($A/\sigma \leq 1$), the Rician distribution is different from Gaussian distribution. However, for larger SNR values ($A/\sigma \geq 3$), the Rician distribution can be approximated by a Gaussian distribution.

A special case of the Rician distribution is obtained in image regions where only noise is present (*A*=0). This case is known as the *Rayleigh* distribution and reduces equation 4.3 to:

$$P_M(M) = \frac{M}{\sigma^2} e^{-\frac{M^2}{2\sigma^2}} \tag{4.4}$$

The mean and variance respectively of this distribution are given by:

$$\bar{M} = \sigma\sqrt{\pi/2} \text{ and } \sigma_M^2 = (2 - \pi/2)\sigma^2$$

This can be used for evaluating the true noise magnitude, $\sigma^2$, from the image. Another limit of the image distribution equation is when the SNR is very large:

$$P_M \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(M-\sqrt{A^2+\sigma^2}\right)^2}{2\sigma^2}} \tag{4.5}$$

This shows that for regions with large signal intensities, the noise distribution can be approximated by a Gaussian with variance $\sigma^2$ and mean $\sqrt{A^2 + \sigma^2}$.

## 4.3 Review of previous denoising approaches

Numerous approaches to denoising MR images have been proposed, including approaches based on anisotropic diffusion (Murase et al., 2001, Gerig et al., 1992), the wavelet transform (Bao and Zhang, 2003), the non-local means algorithm (Buades et al., 2004, Buades et al., 2005b, Buades et al., 2005a), non-parametric neighbourhood statistics/estimation (Awate and Whitaker, 2005b, Awate and Whitaker, 2007, Awate and Whitaker, 2005a, Awate and Whitaker, 2006), and $k$-space filtering (Tisdall and Atkins, 2005). The principal approaches are discussed below together with simple Gaussian low-pass filtering (GLPF) (Gonzalez and Woods, 2002) and bilateral filtering (Tomasi and Manduchi, 1998).

### 4.3.1 Simple Gaussian Filter

One of the simplest techniques for noise removal is the Gaussian low pass filter (GLPF) (Gonzalez and Woods, 2002). It is a smoothing filter and whilst it can effectively eliminate high frequency additive noise, it does so at the expense of blurring fine details and sharp edges in the image. The amount of filtering is directly related to the standard deviation of the Gaussian kernel.

### 4.3.2 Bilateral Filter

A more sophisticated filter is the bilateral filter (Tomasi and Manduchi, 1998), which is based on a combination of two filters, one which is a function of spatial distance, and the other is a function of intensity difference. In the discrete case, the bilateral filter can be

described as follows. Given an image (discrete function), *f*, the filtered value of the image at position *x* is defined by:

$$h(x) = \frac{\sum_{\varepsilon} f(\varepsilon) \cdot c(\varepsilon,x) \cdot S(f(\varepsilon),f(x))}{\sum_{\varepsilon} c(\varepsilon,x) \cdot S(f(\varepsilon),f(x))}, \tag{4.6}$$

where $c(\varepsilon, x)$ measures the distance between the neighbourhood centre, $x$, and a nearby pixel, $\varepsilon$; and $S(f(\varepsilon), f(x))$ measures the intensity similarity between the neighbourhood centre, $x$, and a nearby pixel, $\varepsilon$.

An important case of the bilateral filter is when both similarity measures are Gaussians. In this case $c(\varepsilon, x)$ is defined as:

$$c(\varepsilon, x) = e^{-\frac{1}{2}\left(\frac{||\varepsilon - x||}{\sigma_d}\right)^2}, \tag{4.7}$$

and $S(f(\varepsilon), f(x))$ as:

$$S(\varepsilon, x) = e^{-\frac{1}{2}\left(\frac{||f(\varepsilon) - f(x)||}{\sigma_r}\right)^2}, \tag{4.8}$$

where $\sigma_d$ is the spatial spread and $\sigma_r$ is the photometric spread.

The bilateral filter usually performs better than a linear filter, such as a GLPF, in terms of preserving sharp edges and fine details while still eliminating a significant amount of noise (Tomasi and Manduchi, 1998). It can be easily implemented for 3D images (Jiang et al., 2003), including MRI data (Walker et al., 2006, Xie et al., 2006).

### 4.3.3 **Anisotropic Diffusion filter**

The idea of isotropic diffusion is derived from the heat (isotropic) diffusion process in nature. The basic idea is to embed the original image $I_0(x, y)$ in a series of images, $I(x, y, t)$, obtained by convolving the original image with a Gaussian kernel, $G(x, y; t)$, of variance $t$:

$$I(x,y,t) = I_0(x,y) * G(x,y;t), \tag{4.9}$$

where larger values of $t$ correspond to details at a coarser resolution. This can be viewed as the heat diffusion equation:

$$I_t = div(\nabla I) = \Delta I = I_{xx} + I_{yy} \tag{4.10}$$

with the initial condition $I(x,y,0) = I_0(x,y)$.

As demonstrated by Perona and Malik (Perona and Malik, 1990), the addition of an anisotropic 'diffusion function', $c(x,y,t)$, to the basic heat equation yields an adaptive denoising process:

$I_t = div(c(x,y,t)\nabla I) = c(x,y,t)\Delta I + \nabla \mathrm{c} \cdot \nabla I.$

Ideally, the diffusion function should be chosen such that smoothing is performed on relatively uniform regions but is suppressed on edges. Murase et al. (Murase et al., 2001) propose the following diffusion function, based on the Tukey's bi-weight model (Black et al., 1998), for denoising contrast-enhanced MR images:

$$c(x,y,t) = \begin{cases} \left[1 - \left(\frac{|\nabla I(x,y,t)|}{\sigma}\right)^2\right]^2 & |\nabla I(x,y,t)| \le \sigma \\ 0 & otherwise \end{cases}, \tag{4.11}$$

where σ is a scale parameter. Murase et al. propose that an appropriate value for σ is:

$$\sigma = 1.4826 \cdot MAD(\nabla I), \tag{4.12}$$

where the MAD (median absolute deviation) is given by:

$$MAD(\{x_i\}) = median\{|x_i - median(\{x_i\})|\}. \tag{4.13}$$

Gerig et al. (Gerig et al., 1992) propose the following two alternatives for the diffusion function:

1. $c(x,y,t) = e^{-\left(\frac{|\nabla I(x,y,t)|}{k}\right)^2}$
2. $c(x,y,t) = \frac{1}{1+\left(\frac{|\nabla I(x,y,t)|}{k}\right)^{1+a}}$

where $a>0$ is constant and $k$ is chosen according to the noise level and the edge strength. Denoising based on a higher order differential equation has also been proposed in the literature (Lysaker et al., 2003).

### 4.3.4 **Noise filtering using the wavelet transform**

Over the last twenty years, the wavelet transform has received a great deal of attention in many fields. In the fields of signal and image processing, it is commonly used for denoising. Not surprisingly, several wavelet-based denoising methods have been proposed for MR images (Bao and Zhang, 2003, Nowak, 1999, Wood and Johnson, 1999, Pizurica et al., 2003, Pan et al., 1999). The main advantages of the discrete wavelet transform (DWT) over other image representations, such as the discrete Fourier transform, are that it is a sparse representation that can describe local features, either spatially or spectrally. The sparse representation makes it possible to filter out most of the noise whilst at the same time preserving the high frequency features (fine structures) of the signal.

Nowak (Nowak, 1999) proposed a wavelet-domain filter for denoising MR images specifically designed to handle Rician noise. The method is based on filtering the squared magnitude image (sum of the squared real and imaginary components) computed from the k-space data, and requires that the underlying noise variance be specified or estimated. This approach overcomes the signal-dependent bias that may be introduced, especially in low SNR regions of the image, if the real and imaginary components are filtered separately. Nowak advocates the use of the Haar wavelet to preserve fine detail.

Wood and Johnson (Wood and Johnson, 1999) proposed an MR image denoising method based on the wavelet packet denoising method developed by Coifman and Woog (Coifman and Wickerhauser, 1994, Coifman et al., 1997). The method comprises three elementary steps: wavelet packet transformation, best-basis selection and coefficient thresholding. The method is particularly suited to images that are contaminated by Rician noise with low SNR.

Bao and Zang (Bao and Zhang, 2003) propose a MR image denoising method based on an adaptive multiscale products threshold. The method uses the Mallat and Zhong (Mallat and Zhong, 1992) wavelets family and is based on the evaluation of scale products which are defined as:

$$P_j f(x) = \prod_{i=-k1}^{k2} W_{j+i} \cdot v(x)\,, \tag{4.14}$$

where $k1$ and $k2$ are non-negative integers, and $W_j$ is the wavelet coefficient of scale $j$ of the image $v$ at position $x$. The algorithm consists of the following three steps:

1. Compute the DWT of input image $v$ up to $J$ scales

2. Calculate the multiscale products $P_i v$, select the thresholds $t_p(j)$, and then threshold the wavelet coefficients as follows:

$$\widehat{W}_j v(x) = \begin{cases} W_j v(x) & P_j v(x) \geq t_p(j) \\ 0 & otherwise \end{cases}; \tag{4.15}$$

3. Recover the image from the thresholded wavelet coefficients $W_j v(x)$.

The thresholds, $t_p(j)$, are calculated as a function of the estimated noise in the image in the scale product domain.

### 4.3.5 **Non-local Means (NLM)**

The NLM algorithm was proposed by Buades *et al.* (Buades et al., 2004). The discrete version of the NLM algorithm can be described as follows. Let $I$ be a discrete grid of pixels (voxels), and let $v = \{v(i) | i \in I\}$ be a noisy image. The estimated, denoised, value $NL(v)(i)$ is computed as a weighted average of the image pixels:

$$NL(v)(i) = \sum_{j \in I} w(i,j) v(j), \tag{4.16}$$

where the weight values $\{w(i,j)\}$ depend on the similarity between the pixels $i$ and $j$, and satisfy the conditions $0 \leq w(i,j) \leq 1$ and $\sum_j w(i,j) = 1$.

The definition of the weights $w(i,j)$ relies on the definition of a *neighbourhood* system on $I$ and of a distance measure between two neighbourhoods or *similarity windows.* A neighbourhood system on $I$ is a family $N = \{\mathrm{N}_i\}_{i\in I}$ of subsets of $I$ such that for all $i \in I$:

(i) $i \in N_i$

(ii) $j \in N_i \Rightarrow i \in N_j$

The subset $N_i$ is called the *neighbourhood* or the *similarity window* of $i$. Let the restriction of $v$ to the neighbourhood $N_i$ be denoted by $v(N_i)$, where

$$v(N_i) = \{v(j) \mid j \in N_i\}\,. \tag{4.17}$$

The Gaussian weighted Euclidian distance between two similarity-windows is defined:

$$S(V_1, V_2) = \|V_1 - V_2\|_{2,a}^2, \tag{4.18}$$

where $a$ is the standard deviation of the Gaussian and $V_1$ and $V_2$ represents two vectors that represent the intensity values in the two similarity windows. The weights associated with this distance are defined by:

$$w(i,j) = \frac{1}{z(i)} \cdot e^{-\frac{S\left(v(N_i),v\left(N_j\right)\right)}{h^2}}, \tag{4.19}$$

where $z(i)$ is the normalisation factor $z(i) = \sum_j e^{-S\left(v(N_i),v\left(N_j\right)\right)/h^2}$ and the parameter $h$ controls the decay of the weights and is usually related to the level of noise in the image (Buades et al., 2004, Buades et al., 2005b). Hence, a natural choice for $h$ will be of the form $h = c \cdot \sigma$ where $c$ is a scalar and $\sigma$ is the level of noise in the image.

The NLM algorithm exploits the redundancy of information contained within an image. This redundancy typically exists in high-resolution natural images, because they usually contain textured and smooth regions. However, in the case of medical images, care must be taken, because similar textures in two distant areas of the image may have different meanings (e.g. similar tissue texture from different regions of the brain) and therefore

should not be compared. Several variations on the NLM algorithm have been proposed for denoising 3D MRI (Coupe et al., 2008, Manjon et al., 2007, Manjón et al., 2008, Kervrann et al., 2007), DW-MRI, and DT-MRI data (Wiest-Daessle et al., 2007, Wiest-Daessle et al., 2008).

A drawback of the NLM algorithm is that its time complexity is $O(N \cdot w \cdot s)$ for an image, where $N$ is the number of pixels in the image, $w$ is the number of pixels in a similarity window and $s$ is the number of comparisons for each similarity window (i.e. the search area). For DCE MRI data, this complexity is squared relative to 2D, which often yields non-practical running times. Methods to address this issue include the selective comparison of similarity windows (Mahmoudi and Sapiro, 2005, Coupe et al., 2008) and Laplacian Pyramids (Liu et al., 2008).

### 4.3.6 **Summary**

None of the methods proposed in the literature is specifically designed for DCE-MRI and none actually exploits the redundancy of information that exists in DCE-MRI data. Moreover, some algorithms, such as the NLM, actually seek to exploit redundancy of information that may be false and introduce artefacts in clinical images. This point is discussed further and experimentally demonstrated in Section 4.5.

## 4.4 **Review of methods for evaluating denoising performance**

The evaluation and comparison of different denoising methods is a difficult problem, to date and to the knowledge of the author, no optimal method has been found. Many measures and methods have been proposed in the literature, but comparative results strongly depend on the choice of measurement (Marron and Tsybakov, 1995, Kohler and Lorenz, 2005) and thus depends on the target application. The evaluation of a denoising method usually starts with the creation or acquisition of a noise free image. The noise free image can be synthetically created or acquired in special conditions that produce an approximately noiseless image. Next, random noise is added to the noise free image (according to the 'real life' expected distribution) and the compared denoising algorithms are then applied to the noisy image. The resulting (i.e. denoised) images are then compared to the original (noise free) image using a visual comparison (human eye) and one or more distance measurements (Gonzalez and Woods, 2002).

Numerous measurements can be found in the literature and have been reviewed by Dangeti (Dangeti, 2002). Dangeti also proposes a comparison method that is a combination of visual comparison and a calculation of the SNR in the denoised image. Measuring the SNR of the denoised image, however, may be misleading, because over-smoothed images tend to have a very high SNR. Also SNR does not provide an indication about the visual similarity between the denoised image and the original one.

Maron and Tsybakov (Marron and Tsybakov, 1995), propose a novel method for comparing the original with the denoised signal, using a 'graph representation', which is constructed in the following way.

Let $f$ be the original signal and $f^{den}$ be the denoised signal. Then, the 'symmetrical visual error measure' ($VE$) between the original and the denoised signals will be:

$$VE_2(f, f^{den}) = \left( \int \bar{d}\left( (t, f(t)), G_{f^{den}} \right)^2 dt \right)^{\frac{1}{2}} \tag{4.20}$$

Where $G_{f^{den}}$ is the graph of the function $f^{den}$, and $\bar{d}\big((x, y), A\big)$ is the minimal Euclidean distance from the point $(x, y) \in \mathbb{R}^2$ to a set $A \subset \mathbb{R}^2$. The 'symmetrised version' is

$$SE_2(f, f^{den}) = (VE_2(f, f^{den})^2 + VE_2(f^{den}, f)^2)^{\frac{1}{2}} \tag{4.21}$$

This comparison between images allows more flexibility for small consistent changes in the denoised image, such as a small shift. However, this flexibility also implies that there might be a lack of sensitivity in the method for small artefacts in the denoised image, which makes it not suitable for clinical image evaluation, where small details are often important.

Kohler and Lorenz (Kohler and Lorenz, 2005) compare several denoising methods for one-dimensional signals. Their proposed way for evaluating the denoised signal quality includes Maron and Tsybakov's measurement and a set of distance measurements, including:

The $L_1$ norm:

$$\|f - f^{den}\|_{L_1} = \int |f(t) - f^{den}(t)| dt \tag{4.22}$$

The $L_2$ norm:

$$\|f - f^{den}\|_{L_2} = (\int |f(t) - f^{den}(t)|^2 dt)^{\frac{1}{2}} \tag{4.23}$$

And the $L_\infty$ norm:

$$\|f - f^{den}\|_{L_\infty} = \max_t |f(t) - f^{den}(t)| \tag{4.24}$$

This comparison method produces a set of numbers that represent different sensitivity to outliers in the differences between the images. However, analysing a set of numbers, rather than one, is only a partial solution, because we would like to present the value of the similarity using a single number that can then be easily compared between different candidate denoising methods.

Buades et al. (Buades et al., 2004) review several image denoising methods and compare them. The comparison method, proposed by Buades et al. is based on a visual quality comparison and a mean square error measurement between the original image and the denoised one. One of the visual qualities that Buades et al. propose is the 'method noise' that is the difference image between the noisy (original) image and the denoised version of the image. Buades et al. suggest that a 'method noise' of a good denoising algorithm should look like white noise and present no structures from the original image that were possibly removed by the denoising process. However, in clinical images, different structures have different meanings and contextual information is of great importance in evaluating clinical images visually.

Based on the methods proposed in the literature, it can be concluded that a comparison between the denoising methods must include both quantitative (distance-based measure) evaluation and qualitative (visual) evaluation. In this research we have selected the root-mean-squared-error (RMSE) to be the quantitative measurement, mainly because it is well known and easy to interpret, while the visual evaluation was

selected to be performed on the images (noise free and denoised) by professional assessors that posses either relevant scientific or clinical knowledge.

## 4.5 A new algorithm for denoising DCE-MR data

Recall that DCE-MRI data consists of a set of image volumes acquired before, during and after the injection of a contrast agent. In the absence of any patient movement, it is expected that the volumes will be spatially aligned and have similar voxel values up to the effects of noise and localised contrast enhancement. In practice, of course, the volumes are not perfectly aligned and patient motion will lead to additional variation in the intensity of voxels over time. Nevertheless, it must be the case that the temporal axis contains redundant information, similar to that found in video sequences (Buades et al., 2008). Clearly, in DCE-MRI, a simple averaging over the temporal dimension will not be able to achieve the satisfactory attenuation of noise, the preservation of contrast enhancement and preservation of fine details in the image data. A simple extension of the original NLM algorithm (hereinafter denoted the extended NLM or ENLM) that additionally searches for similarity windows across the temporal dimension is also unsatisfactory for two reasons. First, its similarity metric does not account for local changes in intensity caused by contrast enhancement. Second, while the 'non-local' property of the algorithm means that it can eliminate small differences between different parts of similar tissues because of textural similarity, in the case of MR images, this can lead to the elimination of diagnostically important details. This is especially true when the differences are small or manifest themselves in a relatively small area/volume. Consequently, its suitability and potential for denoising medical images in general is much less than for natural images. Whilst the algorithm may yield very good results in terms of distance-based quality measurements (e.g. MSE) diagnostically-important small details may be lost and possibly false structures introduced. To our knowledge this issue, demonstrated in Figure 4.1, has neither been discussed nor addressed in the literature. To overcome the first drawback, what is needed is a similarity metric that takes into account local enhancement. This leads to the proposed variation on the NLM algorithm, Dynamic Non-Local Means (DNLM). Specifically, we redefine the similarity metric to be:

$$\bar{S}(V_1, V_2) = S(V_1, C(V_1, V_2) \cdot V_2) = \|V_1 - C(V_1, V_2) \cdot V_2\|_{2,a}^2 \tag{4.25}$$

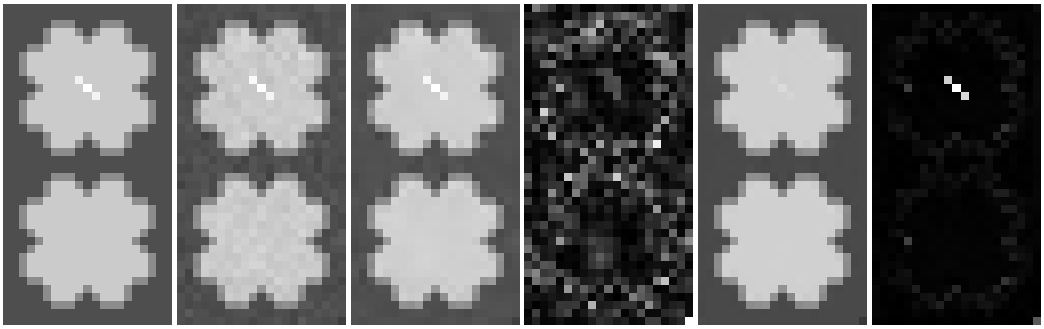
and the definition of weights to be:

$$\bar{w}(i,j) = \frac{C(v(N_i), v(N_j))}{z(i)} \cdot e^{-\frac{\bar{S}(v(N_i), v(N_j))}{h^2}} \tag{4.26}$$

where

$$C(V_1, V_2) = \begin{cases} \frac{E(V_1)}{E(V_2)} & t_1 \neq t_2 \ \ and \ |E(V_1) - E(V_2)| > \sigma \\ 1 & otherwise \end{cases}, \tag{4.27}$$

where $t_1$ and $t_2$ are the temporal components from which the vectors $V_1$ and $V_2$ were selected, σ is the (estimated) level of noise in the image and $E(V)$ denotes expected value. The new weights (equation 4.26) are to replace the original ones (equation 4.19). In choosing the level of noise, $\sigma$, to be the threshold for the normalisation, the underlying assumption is that changes in signal that are smaller than the level of noise are likely to be related to noise, while larger changes indicate a real change in signal (and thus should not be normalised). In making this change, it is assumed that different pixels in two similarity windows from different time points in the acquisition will enhance in the same manner if they are part of a similar tissue. To address the second drawback (the propensity of the 'non-local' property to eliminate small details), the search for similarity windows must be limited to a relatively small area in the spatial domain. The search over the temporal domain, however, is not limited. In other words, the algorithm searches for 'dynamic' similarity windows along the whole temporal axis. Therefore, the DNLM algorithm makes it possible to filter the image in the spatial domain, while exploiting the similarity over the temporal axis without changing the effect of contrast enhancement or blurring the image in the event of patient movement. An additional advantage of the restricted similarity window search is that the execution time of DNLM is significantly faster than the ENLM.



(a) (b) (c) (d) (e) (f)

Figure 4.1: The disadvantage of the 'non-local' property of NLM in DCE MRI
(a) Synthetic, noiseless, image of two similar tissues, the upper one with a potential abnormality (e.g. enhancing tissue), (b) image (a) after the addition of Rician noise to yield an SNR of 28db, (c) image (b) after the application of NLM with a search distance of 1 pixel, yielding an MSE of 8.75, (d) the 'method noise' (Buades et al., 2005b) of (c), (e) image (b) after the application of the NLM with a search distance of 20 pixels, yielding a 'better' MSE of 5.33, (f) the 'method noise' of (d). It can be clearly seen in both the denoised images and the 'method noise' images that, although the larger search distance yields a smaller MSE, important detail is lost.

## 4.6 **Empirical evaluation**

To evaluate the performance of the DNLM algorithm, we elected to compare its denoising ability to that of seven other algorithms: Gaussian low pass filter (GLPF), the bilateral filter anisotropic diffusion (AD) (Murase et al., 2001), wavelet adaptive multiscale products threshold (WAMPT) (Bao and Zhang, 2003), traditional wavelet thresholding (TWT) (Pan et al., 1999, Bao and Zhang, 2003), NLM, and the ENLM (the trivial temporal extension of the NLM discussed in section 4.5). The GLPF was chosen as a reference filtering approach. The others were chosen because they represent several different approaches to the MR image denoising problem. Three experiments were performed.

Experiment 1: A quantitative evaluation using an artificially-generated DCE-MRI image sequence.

Experiment 2: A quantitative evaluation using real clinical DCE-MRI data.

Experiment 3: A qualitative evaluation using real clinical DCE MRI data.

The purpose of the quantitative experiments was to evaluate the denoising performance purely in terms of the MSE. However, as noted in Section 4.5, it is possible for an algorithm to yield results that appear to be good both visually and in terms of the MSE and yet, on closer inspection, contain geometric artefacts. This was the rationale for

undertaking the qualitative experiment. In all experiments, the denoising algorithms were applied to individual sagittal slices rather than to entire volumes. This was done because, in the clinical data, the slice pixels were square (~0.8mm) with a much smaller dimension than the slice thicknesses (~5mm).

The DCE-MRI data used directly in these experiments, hereinafter called the *test data*, originate from 16 routine clinical breast MRI examinations performed by Queensland X-Ray (QXR) and four breast MRI examinations of subjects with diagnosed breast cancer performed at the Wesley Research Institute (WRI). The QXR DCE-MRI data were acquired with fat-suppression and the WRI data without. Several additional QXR data sets, hereinafter called the *training data*, were used indirectly to, (i) gauge the range of noise levels that might typically be encountered in DCE MR images acquired in clinical practice to optimise parameter values for the denoising algorithms subsequently evaluated in the experiments (the parameters were optimised to yield the minimum MSE on the training data) and, (ii) to build the synthetic image sequence used in Experiment 1. The approval of the Human Research Ethics Committee of the University of Queensland was obtained for this study.

The specific instantiations of the denoising algorithms evaluated in the experiments are as follows. For the GLPF, a 2D Gaussian with a standard deviation of one pixel was used. For the Bilateral filter, Gaussian similarity measures were used with $\sigma_d$ (spatial spread) equal to 1.8 pixels and $\sigma_r$ (photometric spread) equal to 4 times the amount of the estimated noise. For the NLM, a 11×11 pixel similarity window was used (corresponding to approximately 9×9 mm) and $h$ was taken to be $1.75\sigma$ (where σ is the estimated level of noise in the image). For the ENLM, a 7×7 pixel similarity window was used (approximately 5.5×5.5 mm) with $h = 1.25\sigma$ and for the DNLM, a 5×5 similarity window was used (approximately 4×4 mm) with $h = 1.5\sigma$. The search area for the NLM algorithm was defined to be one pixel in each spatial direction, whilst for the DNLM and ENLM algorithms, it was defined to be one pixel in each spatial direction and all time points in the temporal dimension. The constraint on the spatial search area was imposed to minimise the drawback of the 'non-local' property of the NLM algorithm (discussed in Section 4.5) while still accounting for some patient movement. The anisotropic diffusion parameter was chosen to be five times the amount of estimated noise. For the WAMPT algorithm, $k1$=1 and $k2$=0 were used (these were the default

values in the implementation made available to us by Bao & Zhang (Bao and Zhang, 2003). In the case of the TWT algorithm no parameter values needed to be selected a priori.

The median-absolute-value (MAV) (Bao and Zhang, 2003) estimator, which is based on Donoho's noise estimator (Donoho and Johnstone, 1994), was chosen as the noise estimator for the experiments, based on experimental evaluation. The estimator was tested using three real images from the training data (selected because they contained very little visual evidence of noise). Several different levels of simulated Rician noise were imposed (Manjón et al., 2008) on these images, as shown in the first column of Table 4.1. The MAV estimator was found to yield biased, although linearly-related, estimates of the noise level. A line was then fitted to the estimated results to correct for the bias and an unbiased estimator of the noise was thus determined to be

$$\sigma_c = 1.25\sigma - 6.$$

Table 4.1: Estimating Rician noise using the MAV estimator

| Imposed noise level | SNR (db) | Estimated Noise ($\sigma_c$) |
|---|---|---|
| 10 | 32.9 | 9.6 |
| 40 | 20.9 | 40.7 |
| 80 | 14.9 | 80.6 |
| 120 | 11.3 | 117.0 |
| 160 | 8.8 | 156.2 |

### 4.6.1 **Experiment 1: Quantitative evaluation using an artificially generated DCE-MRI image sequence**

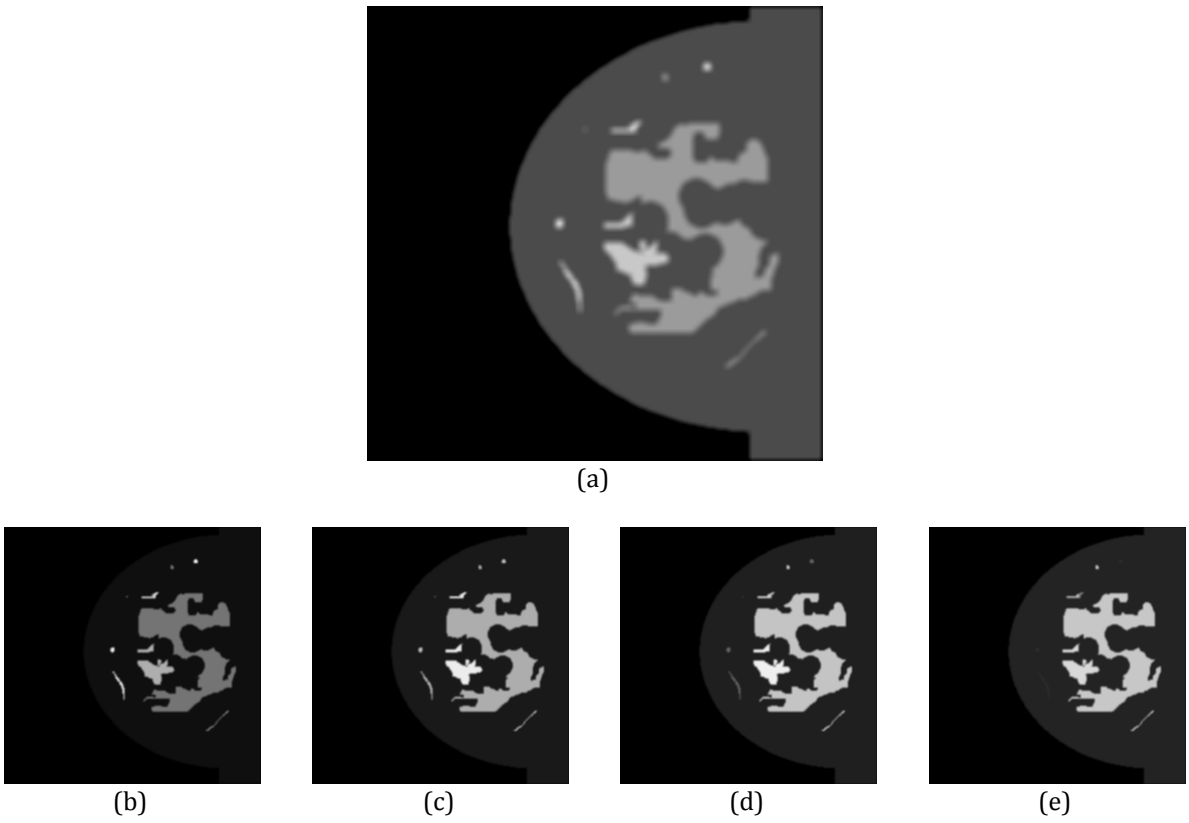
The aim of this experiment was to quantitatively evaluate the performance of each denoising algorithm using an artificially-created DCE-MRI sequence to which known levels of noise could be added. The artificial DCE-MRI sequence was created from the synthetic image shown in Figure 4.2. The image resembles a sagittal DCE-MRI image of a breast. It contains several structures that mimic different tissue types. The intensities of the individual structures were independently varied to yield an artificial temporal sequence. Realistic intensity variation over time was achieved by superimposing enhancement variations derived from a real breast DCE-MRI data set selected from the

training data. More specifically, a parametric model of contrast enhancement (Gal et al., 2007b) was fitted voxel-wise to the selected data set and then used to generate plausible enhancements in the synthetic temporal sequence. The resulting artificial data set comprised 7 slice images of 256×256 pixels. For the enhancement model, we used the following parameter values, (i) *a=6.2, b=1.4, c=1.2* to yield a curve shape which is typical for malignant enhancement and, (ii) *a=1.2, b=3.8, c=0.8* to yield a curve shape that is typical for benign enhancement. Several intermediate curve shapes were also used. A GLPF, with a standard deviation of 2 pixels, was then applied to the synthetic image sequence to smooth the artificially-created step edges. Next, three separate noisy image sequences were generated from this data set. Each was obtained by adding a specific level of simulated Rician noise (Manjón et al., 2008) to the synthetic image sequence. In two cases, the level of noise added was consistent with that measured in the training data (SNR levels of about 19.8 and 13.8 db), and in the third, a higher level of noise was used (SNR of about 10.2 db). Next, each of the denoising algorithms was independently applied to the third image in each of the noisy sequences. The third image was chosen because, in the real DCE-MRI data set, most enhancing tissues showed significant enhancement at that time point. The result from each algorithm was then compared to the noiseless version of this image using the mean-squared-error (MSE) metric. The results are shown in Table 4.2. In all cases, the DNLM algorithm yielded the smallest MSE.

### 4.6.2 **Experiment 2: Quantitative evaluation using real DCE-MRI data**

The aim of this experiment was to quantitatively evaluate the performance of each denoising algorithm using real DCE-MRI data. One post-contrast slice image was randomly selected from each of the 20 different DCE-MRI data sets in the test data. Rician noise was then added to these images and to their counterparts over time to yield 20 noisy image sequences. The level of noise added was chosen to yield SNRs consistent with the worst SNR estimated from the training data (about 8 db). Next, each of the algorithms was independently applied to the noise-corrupted versions of the selected post-contrast slice images and each of the denoised images was then corrected for Rician noise bias viz. (Manjón et al., 2008)

$$\bar{f}(x,y) = \sqrt{\left[\hat{f}(x,y)\right]^2 - 2\sigma^2} \tag{4.28}$$



(a)

(b) (c) (d) (e)

Figure 4.2: Synthetic breast DCE-MRI image used in Experiment 1 to generate the artificial DCE-MRI image sequences
Image (a) demonstrates the enhancement coefficients of different tissues. The different shapes in the image represent different tissue types with different enhancement curves. Images (b)-(e) are the substruction of the pre-contrast from the first, second, third and forth post-contrast images, respectively. The images are presented after the blurring stage.

Where $\bar{f}(x,y)$ is the bias-corrected version of the denoised image $\hat{f}(x,y)$. Each resultant image was then compared to the corresponding original image using the MSE metric. The MSE results are summarised in Table 4.3 and shown graphically, using box plots, in Figure 4.4. The results show that the median and mean MSE for the DNLM algorithm is smaller than that for any other method and that the spread of the MSE values about the median (interquartile range) is also small. The Friedman test was used to test the null hypothesis that all of the algorithms perform equally well, versus the alternate hypothesis that they do not. The different denoising methods were taken to be

the column effects, with the different images being taken to be the row effects. The test was performed using the Statistics Toolbox in MATLAB (2007). The p-value for the test was smaller than $10^{-4}$, thus providing evidence, at the $\alpha$=0.05 level of significance, that the algorithms do not all perform equally well. A post hoc multiple-comparison test consisting of seven paired-sample Wilcoxon tests was also performed. More specifically, the performance of each denoising method was compared in turn to that of the DNLM method using a paired-sample Wilcoxon test. In each test, the null hypothesis was that the DNLM method and the other method perform equally well, versus the alternative hypothesis that the DNLM method performs better; i.e. yields a smaller MSE. The unadjusted p-values for the individual paired-sample tests are shown in Table 4.4. The multiple comparison test was performed at the $\alpha$=0.05 level of significance dictating that, using the Bonferroni correction, each pair-wise test be performed at the $\alpha$=0.05/7=0.0071 level of significance. Thus, it can be concluded, at the $\alpha$=0.05 level of significance, that the DNLM algorithm performs better than all of the other denoising algorithms.

Table 4.2: Results of the quantitative evaluation using three artificially generated noisy DCE-MRI image sequences (Experiment 1)
The numbers in the table are the MSE between the bias-corrected denoised image and the corresponding (noiseless) artificial image (lowest MSE is highlighted in bold).

| Method | Sequence 1 (SNR 19.8db) | Sequence 2 (SNR 13.8db) | Sequence 3 (SNR 10.2db) |
|---|---|---|---|
| AD | 1109 | 3989 | 8443 |
| WAMPT | 1522 | 3795 | 7606 |
| DNLM | **983** | **3487** | **7535** |
| GLPF | 5858 | 8423 | 12681 |
| Bilateral | 5864 | 8352 | 12066 |
| TWT | 1609 | 4159 | 8170 |
| NLM | 1244 | 4452 | 9532 |
| ENLM | 1199 | 4168 | 8799 |

Table 4.3: Means and five number summaries of the MSE values for all the methods in Experiment 2 (Minimum, mean, median, interquartile range (IQR) and maximum value), lowest values are highlighted in bold

| Method | Min | MEAN | MEDIAN | IQR | Max |
|---|---|---|---|---|---|
| AD | 946 | 3773 | 1475 | 2412 | 24046 |
| WAMPT | 815 | 3578 | 1330 | 2322 | 23971 |
| DNLM | **542** | **1584** | **749** | 882 | **8540** |
| GLPF | 647 | 2072 | 1023 | 1729 | 9494 |
| Bilateral | 796 | 3783 | 1321 | 2446 | 25183 |
| TWT | 883 | 3715 | 1376 | 2390 | 24526 |
| NLM | 692 | 1967 | 924 | 1206 | 9664 |
| ENLM | 545 | 1669 | 788 | **878** | 8987 |

Table 4.4: Unadjusted p-values for the post hoc multiple comparison test of Experiment 2

| Test No. | Null hypothesis | Alternate hypothesis | p-value |
|---|---|---|---|
| 1 | $\mu_{\text{DNLM}} = \mu_{\text{NLM}}$ | $\mu_{\text{DNLM}} < \mu_{\text{NLM}}$ | $< 10^{-6}$ |
| 2 | $\mu_{\text{DNLM}} = \mu_{\text{ENLM}}$ | $\mu_{\text{DNLM}} < \mu_{\text{ENLM}}$ | $< 10^{-4}$ |
| 3 | $\mu_{\text{DNLM}} = \mu_{\text{Bilateral}}$ | $\mu_{\text{DNLM}} < \mu_{\text{Bilateral}}$ | $< 10^{-6}$ |
| 4 | $\mu_{\text{DNLM}} = \mu_{\text{TWT}}$ | $\mu_{\text{DNLM}} < \mu_{\text{TWT}}$ | $< 10^{-6}$ |
| 5 | $\mu_{\text{DNLM}} = \mu_{\text{GLPF}}$ | $\mu_{\text{DNLM}} < \mu_{\text{GLPF}}$ | $< 10^{-6}$ |
| 6 | $\mu_{\text{DNLM}} = \mu_{\text{WAMPT}}$ | $\mu_{\text{DNLM}} < \mu_{\text{WAMPT}}$ | $< 10^{-6}$ |
| 7 | $\mu_{\text{DNLM}} = \mu_{\text{AD}}$ | $\mu_{\text{DNLM}} < \mu_{\text{AD}}$ | $< 10^{-6}$ |

### 4.6.3 **Experiment 3: Qualitative evaluation, by expert observers, using real DCE MRI data**

Experiments 1 and 2 yielded quantitative evaluations of the performance of each denoising algorithm. However, the results may be biased because of the nature of the image data used in the experiments and the choice of performance metric (MSE). Also, in Experiment 1, the synthetic image data used may be too simplistic, e.g. the bias field, motion artefacts, and the true complexity of the tissue types and structure are not modelled. In Experiment 2, although real data is used, the reality is that it is not possible to acquire perfectly noiseless data and thus the ground truth may be biased. To provide a more balanced picture it was decided to undertake a subjective (i.e. visual) comparison in Experiment 3. The aim of Experiment 3 was to qualitatively compare the performance of the five best denoising methods from Experiment 2 — the DNLM, ENLM, NLM, WAMPT and GLPF — using real DCE-MRI data. The reason for choosing the best five was to simplify the ranking task (described in detail below), for the expert observers. The data used in Experiment 2 was used again in Experiment 3, i.e. 20 sets of images, each set containing the original, original with added Rician noise, and the

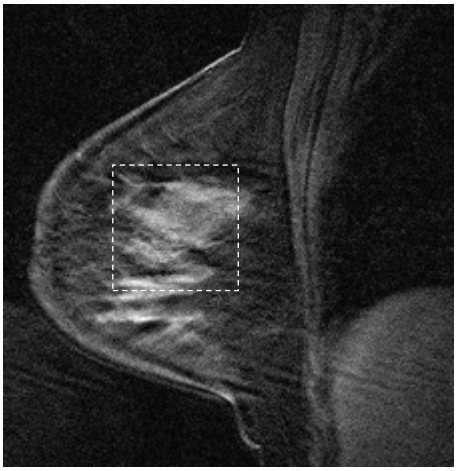
denoised versions of the noisy image after bias correction. A MATLAB (2007) program was written to:

1. present to the user one of the 20 sets of images (original image, original image with added noise, five bias-corrected denoised versions of the noisy image) at a time, in a random order
2. allow the user to arrange (rank) the five bias-corrected denoised images from best (a rank of 5) to worst (a rank of 1)
3. record the ranking.

Each invocation of the program yields the 20 sets of images in random order, and for each set of images, the bias-corrected denoised images are also presented in random order (a sample image is shown in Figure 4.3). A total of 24 expert observers were recruited for the experiment, 14 image/signal-processing experts, and 10 breast MRI radiographers. For practical reasons, the signal-processing observers and the clinical observers undertook the experiment at physically-different locations, using physically-different hardware. However, the lighting and viewing conditions for both groups were matched as closely as possible. In particular, TFT LCD screens were used, under regular office ambient lighting, with a one-to-one mapping of the image pixels to the screen pixels. In addition, the MATLAB program permitted the user to view each image in a 2:1 pixel-doubled format (no interpolation) and, optionally, as a photographic negative (negative grey-scale). Each observer was asked to use the program to rank the denoised images in each set of images from best to worst in terms of 'noise attenuation, the restoration/preservation of fine details and the minimal introduction of artefacts'. Each was instructed to make his/her own decision about the relative importance of each of these quality factors. Each observer completed this task in isolation from the other observers and was blind to the results of the other observers. A summary of the results by method is presented in Table 4.5 and shown graphically using box plots in Figure 4.5. The results show that the median and mean ranks for the DNLM algorithm are better than that for any other method. The different observers were not separated into their professional groups, because a previous experiment showed that both groups of experts rank the images in a similar manner (Gal et al., 2008).

The Friedman test was again used to test the null hypothesis that all of the algorithms perform equally well, versus the alternate hypothesis that they do not. The different

denoising methods were taken to be the column effects, the different observers were taken to be the row effects, and the individual images to be the replicates. The test was performed using the Statistics Toolbox in MATLAB (2007). The p-value for the test was smaller than $10^{-4}$ thus providing evidence, at the $\alpha$=0.05 level of significance, that the algorithms do not all perform equally well. A post hoc multiple-comparison test consisting of four paired-sample Wilcoxon tests was also performed. More specifically, each denoising method was compared in turn to the DNLM method using a paired-sample Wilcoxon test. In each test, the null hypothesis was that the DNLM method and the other method perform equally well, versus the alternative hypothesis that the DNLM method performs better, i.e. that the assigned score (rank) is larger. The p-value for each test was no larger than 0.00410. The multiple comparison test was performed at the $\alpha$=0.05 level of significance so that, using the Bonferroni correction, the level of significance actually used in each paired test was $\alpha=0.05/4\approx0.0125$. Thus, it can be concluded, at the $\alpha$=0.05 level of significance, that the performance of the DNLM algorithm is better than all of the other algorithms.



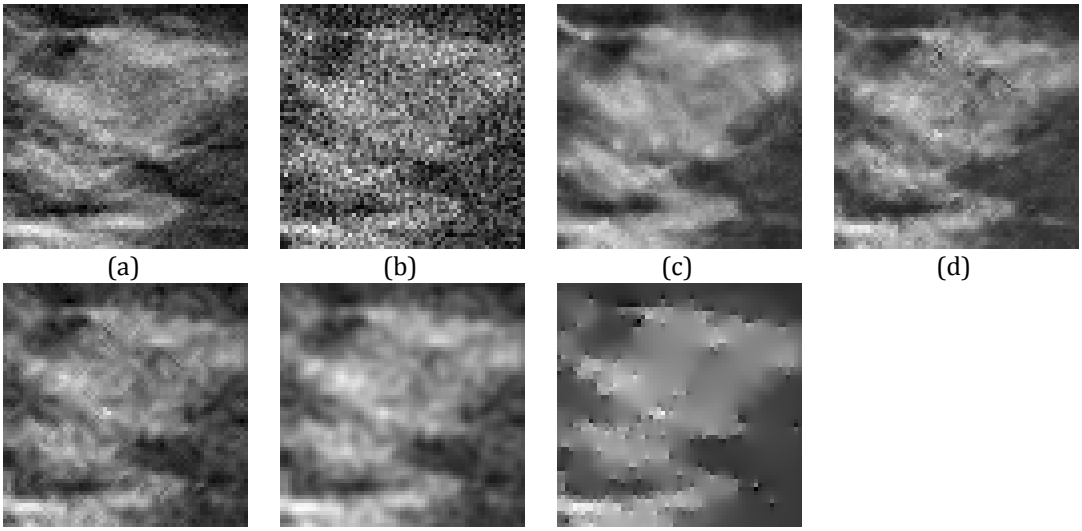



Figure 4.3: Clinical breast DCE-MRI image used in experiment 3. The denoising results are demonstrated on a selected patch of the image.
Top: Original, image, (a) patch from the original image, (b) patch with added noise, (c) denoising result of the DNLM (d) denoising result of the ENLM (e) denoising result of the 2D NLM, (f) denoising result of GLPF and, (g) denoising result of the WAMPT.

Table 4.5: Means and five number summaries of the ranks assigned by all observers in Experiment 3 (Minimum, mean, median, interquartile range (IQR) and maximum value for each of the denoising methods, 1 is worst and 5 is best)

| Method | Min | MEAN | median | IQR | Max |
|---|---|---|---|---|---|
| DLNM | 2 | 4.5 | 5 | 1 | 5 |
| ENLM | 2 | 4.3 | 4 | 1 | 5 |
| NLM | 1 | 2.8 | 3 | 0 | 5 |
| GLPF | 1 | 2.2 | 2 | 0 | 5 |
| WAMPT | 1 | 1.1 | 1 | 0 | 4 |

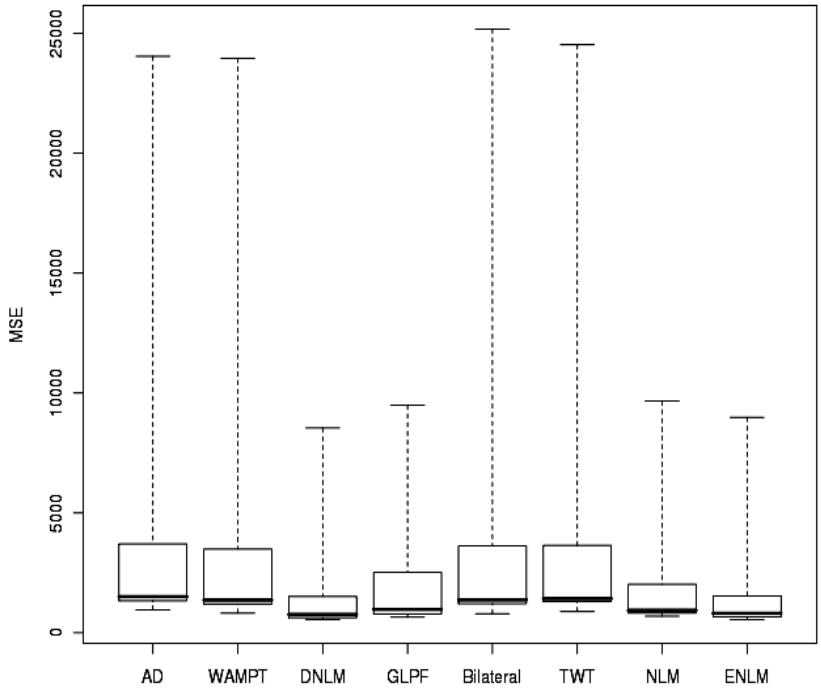



Figure 4.4: Box plots of the MSE values for all the methods in Experiment 2

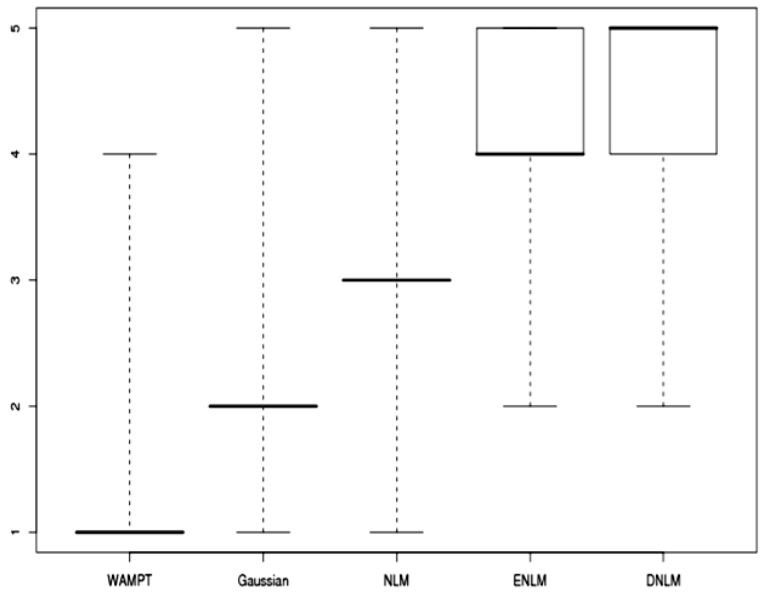



Figure 4.5: Box plots of the ranks assigned by the observers in Experiment 3 to each of the denoising methods (1 is worst and 5 is best)

### 4.6.4 **Discussion and conclusions of empirical evaluation**

Section 4.6 presented details of the empirical evaluations of the performance of the DNLM relative to seven other denoising methods, Gaussian low-pass filtering (GLPF), the original NLM algorithm, a trivial temporal extension of the NLM (ENLM), bilateral filtering, anisotropic diffusion, wavelet adaptive multiscale products threshold, and traditional wavelet thresholding. In particular, three experiments were reported. In the

first experiment, the algorithms were quantitatively evaluated using three synthetically-generated noisy DCE MRI datasets. The results show that for all three data sets, the DNLM algorithm yielded the smallest MSE between the bias-corrected denoised image and its corresponding original noiseless version. In the second experiment, the algorithms were quantitatively evaluated using 20 DCE-MRI data sets from routine clinical breast MRI examinations to which a known level of Rician noise was added. A Friedman test provided evidence, at $\alpha=0.05$ level of significance, that in terms of MSE, the algorithms do not all perform the same. A post hoc multiple comparison test, with Bonferroni correction, provided evidence, at $\alpha=0.05$ level of significance, that the DNLM algorithm yields the smallest MSE between the bias-corrected denoised image and its corresponding original noiseless version. In the final experiment, the algorithms were qualitatively evaluated by 24 expert observers using the 20 data sets from Experiment 2. The qualitative evaluation involved each observer independently ranking the bias-corrected denoised images produced by each algorithm, from worst to best (visual comparison), for each of the 20 data sets. A Friedman test provided evidence, at $\alpha=0.05$ level of significance, that expert observers do not visually perceive each of the algorithms to perform equally well. A post hoc multiple comparison test, with Bonferroni correction, provided evidence, at $\alpha=0.05$ level of significance, that expert observers find the DNLM algorithm to be the best denoising method. Collectively, the qualitative and quantitative results suggest that the DNLM algorithm more effectively attenuates noise in DCE MR images than any of the previously-proposed algorithms. The results suggest that the DNLM algorithm is useful both as a tool for improving image quality for a given resolution and to improve image resolution without compromising the image quality in DCE MRI data. This, in turn, may supply the clinician with more accurate image data for interpretation in addition to providing a better starting point for automatic segmentation and classification by computer.

## 4.7 **Summary**

This chapter has reviewed existing methods for denoising MR and DCE-MR images and methods for comparing denoising algorithms. The chapter introduced a new method, called dynamic non-local means (DNLM), for denoising DCE-MR images. This method exploits the redundancy of information in different volumes of the data acquired at different time points. Also, quantitative (both on synthetic and real data) and

qualitative evaluations of the new method against seven other methods had been introduced. Denoising DCE-MR images improves the quality of the image and paves the way for higher-level processing methods that are discussed in Chapter 5 (segmentation of suspicious lesions) and Chapter 6 (feature extraction and classification of lesions).

# 5. Automatic Segmentation of Enhancing Breast Tissue in DCE-MRI

The amount of data that needs to be examined by a radiologist in DCE-MRI to locate suspicious lesions is huge (e.g. 512x512 voxels x 50 slices x 5 volumes). The automatic segmentation of suspicious lesions in DCE-MRI of the breast is thus attractive, because it may dramatically reduce the amount of data that needs to be examined and draws the radiologist's attention to volumes that have a high probability of containing suspicious lesions. However, while, segmenting the image into the relevant objects and background parts is a crucial step, "it is also, in many cases, one of the more difficult tasks" (Rodenacker and Bengtsson, 2003).

This chapter introduces the challenge of automatic segmentation. Also, it presents a method for the automatic segmentation of enhancing breast tissue. The method is based on seeded region growing and merging. The criteria for growing and merging are based on both the original image intensity values and the fitted parameters of an empiric parametric model of contrast enhancement. The results for the application of the method are also presented on 24 DCE-MRI breast data sets originating from routine clinical breast MRI examinations. The data includes 10 cases of benign enhancement and 14 cases of malignant enhancement (the latter confirmed by histopathology). The results show that the segmentation method has 100% sensitivity for the detection of suspicious regions independently identified by a radiologist. The results suggest that the method has the potential both as a tool to assist the clinician with the task of locating suspicious tissue and as a means for generating quantitative features for the automatic classification of suspicious regions. The core material of this chapter was

presented at 'Digital Image Computing: Techniques and Applications', 2007, Adelaide, Australia (Gal et al., 2007a).

## 5.1 Introduction

"Image segmentation is the process of separating objects from background" (Snyder and Qi, 2004). The segmentation of an image is the partitioning of an image into a set of connected regions, where each region is homogeneous in some sense (e.g. intensity or texture) and is identified by a unique label (Snyder and Qi, 2004). The basic assumption is that the object in the image differs from the background (i.e. everything that is not part of the object) in some properties (e.g. shape, intensity, texture). The result of a segmentation method is usually a list of equivalence classes where each class represents an object or the background.

Classification of objects (e.g. lesions) in the spatial domain is commonly based on the segmentation and different properties of the image, such as morphometric (i.e. shape, size), radiometric (i.e. gray level, histogram) and textural properties. The first step in object classification is usually the segmentation of the object of interest in the image. Robust segmentation is difficult to achieve; thus, the classification process is often expected to overcome the noise and bias that may be introduced by the segmentation step.

Manual segmentation is subjective, and given the vast quantity of data to be analysed in a DCE-MRI data set, the possibility exists that diagnostically-significant regions of enhancement may be overlooked. However, automatic segmentation is challenging, because the temporal and spatial distributions of the contrast agent in suspicious tissue can be highly varied, both for an individual patient and between patients.

## 5.2 Segmentation methods for greyscale images

In this section, an overview of segmentation methods for greyscale images is presented. These methods provide the basis for MR images segmentation techniques. In general, segmentation techniques can be roughly divided into two types of algorithms: those that find the edges and contours in the image, assuming that once a contour is found, the inner side of the contour will be the object while the outer region is the background; and those that define a criteria of membership for a region and search for connected

sets of pixels that satisfy these criteria. Some of these are based on global and local thresholding, which usually involves the analysis of the image histogram (Arifin and Asano, 2006, Abutableb, 1989, Wong and Sahoo, 1989, Otsu, 1979, Cheng et al., 2000, Sezgin and Sankur, 2004). Such techniques attempt to estimate a threshold or thresholds that can create a binary differentiation between the object's intensity levels and the background intensity levels. Global, single value, thresholding techniques are usually not practical for real world images, because of the variety of objects in the image and the presence of noise and different bias effects (e.g. illumination, shadows, optical effects, bias field in MRI). Local thresholding techniques tend to overcome this limitation by dividing the image into small segments or regions of interest and applying a different threshold to each one of them. The size of these segments is usually decided prior to the segmentation process, because they have a major influence on the outcome. Global, multiple threshold binarization techniques, on the other hand, try to manipulate the image histogram by dividing it into intensity subsets or by projecting it into a different space such that the binarization will incorporate additional information, rather than just the intensity histogram (Abutableb, 1989, Cheng et al., 2000, Sezgin and Sankur, 2004). Of these, the entropy-based binarization (Abutableb, 1989) has shown to be both robust and parameter free. The entropy thresholding method is described in detail in Appendix A.

Edge detection techniques attempt to mark the boundaries of the desired object(s) and to segment it/them by identifying the image pixels that lie inside those boundaries. Edge detection methods usually assume meaningful changes of intensity levels or intensity behaviour between different objects in the image (Pal and Pal, 1993). These techniques sometimes approximate the image by a set of planes, manifolds or curves that roughly present the expected behaviour of the intensity as if the image was continuous and noise free.

Inside the edge detector algorithm family, both parallel and sequential edge detectors can be found. Examples of parallel edge detectors are Canny, Prewitt and Sobel edge detectors (Gonzalez and Woods, 2002). Sequential edge detectors tend to solve regularized 'shortest path' type problems that are an outcome of representing the image as an edge-weighted connected graph, where the edge weights are usually a function of the Euclidian distance between pixels, intensity difference, and so on. Some of the most

common examples for such edge detectors are Snakes and Balloons. The main disadvantage of such methods is their dependency on the selection of the starting point and the regularization function for the algorithm (Kass et al., 1988).

An expansion of the dynamic contour edge detectors are the level set/fast marching algorithms. The level set algorithms approximate the equations of motion for a propagating boundary by transforming them into a partial differential equation whose unique solution is the position of the boundary. The fast marching algorithm is a special case of the level set algorithms family with a boundary moving at a speed of $F = F(x, y)$ where F is always positive or negative. This family of algorithms assumes a starting point that belongs to the desired object. Level set algorithms handle corners and cusps naturally, in addition to topological changes. Nevertheless, level set methods assume a pre-known behaviour of the intensity levels inside and on the boundaries of the object, which is incorporated into the formulated speed/time functions of the propagating boundary (Sethian, 1999, Malladi and Sethian, 1996, Malladi and Sethian, 1995, Sethian, 1996).

Another family of methods that uses a similar concept to the fast marching is the seeded region growing (SRG). This family of methods usually requires starting points or 'seeds' from which the growing process will start (Adams and Bischof, 1994, Petrick et al., 1999, Mehnert and Jackway, 1997, Revol and Jourlin, 1997, Xuan et al., 1995, Pohle and Toennies, 2001, Kupinski and Giger, 1998). The attractiveness of the process is that it is fairly intuitive (one can estimate the final result by looking at the image) and that it permits integrating prior information into the algorithm. Suspicious lesions in DCE-MRI of the breast usually consist of a single tissue type (per lesion). Thus, they appear as one (or several) connected component of pixels with similar properties (e.g. textural, kinetic). This has made the SRG algorithm a good candidate for segmenting suspicious lesions in the breast, once a lesion area has been located, by using a pre-processing (e.g. too sensitive segmentation) step.

Other families of segmentation algorithms include probalistic model estimations, such as Markov random fields and the Maximum Likelihood (Pal and Pal, 1993, Rimey and Cohen, 1988). The latter assumes that the data in the image of any one object are statistically independent of the others. It then tries to find a segmentation that has the maximum probability, given the image, assuming that the similarity in the intensity

properties of the neighbouring pixels is proportional to the probability that they belong to the same object.

The robustness of the entropy-based binarization and the basic assumption upon which SRG is based, have made these methods the natural choice in this research. This was the main motivation for using them as the main steps in the segmentation algorithm proposed in this chapter.

## 5.3 Review of breast lesion segmentation techniques

Breast medical images can be acquired in several ways. Among which are the Ultrasound, X-ray mammograms, Computed Tomography (CT), Nuclear Medicine and MRI (including DCE MRI). Each modality yields a slightly different image that presents different properties of the same object. However, all acquisition results are translated to a discrete image that represents the local average intensity of the observed property of the tissue in each pixel/voxel in the image.

When trying to detect malignant tissue in a breast volume, it is assumed that malignant tissues have different characteristics to benign tissue, in the scale of the acquisition intensity results. The differentiation can appear either in rough intensity, in boundary shape, texture or any combination of them. In dynamic imaging, differences can also be observed on the time axis.

One of the common segmentation methods is the SRG. Variations of this method are often used in the transformation of the information extracted from the image, such as the Density Weighted Contrast Enhancement (DWCE) (Petrick et al., 1999). In other cases, strong assumptions, such as the convexity of the lesion (Kupinski and Giger, 1998, Pohle and Toennies, 2001) or the homogeneity of either the intensity levels (Pohle and Toennies, 2001) or the contrast enhancement over time (Lucas-Quesada et al., 2005), are made to simplify the process. Moreover, SRG-based methods in the literature often rely on manual seed selection, which requires the user to review the data and identify the regions of interest. Alternatively, the user is required to mark a region of interest around the lesion to bind the area in which the segmentation is performed (Chen et al., 2006b, Chen et al., 2006a) or even perform a ‘rough’ prior delineation of the lesion that will later be refined by the algorithm (Gilhuijs et al., 1999).

Fully-automatic methods have been also proposed (Hayton, 1998, Li et al., 1995). These, however, usually involve a set of parameters that need to be empirically tuned, and may not fit every dataset (Hayton, 1998) or tend to be highly sensitive to the structure of the data (i.e. perform differently on different datasets) or to the order in which the pixels are scanned (Li et al., 1995). Li et al. also comment in their paper that the algorithm is not generally sensitive to masses.

A fully-automatic segmentation algorithm with a high sensitivity to suspicious lesions is thus desirable. In this chapter, a fully-automatic segmentation algorithm for suspicious lesions in DCE-MRI of the breast is presented. The algorithm demonstrated a sensitivity of nearly 100% to enhancing lesions on DCE-MRI data from clinical practice.

## 5.4 New algorithm for segmenting enhancing lesions in DCE-MRI

The proposed segmentation algorithm makes use of the kinetic model proposed in section 3.3. The proposed segmentation algorithm is shown in Algorithm 5.1. The individual steps of the algorithm are described below.

### 5.4.1 Computing the 'Critical Points Map' (CPM)

The proposed model of enhancement can be described as follows:

$$f(t) = a \cdot t \cdot e^{-\frac{t^c}{b}} \tag{5.1}$$

where $t$ is time, and $a$, $b$, and $c$ are the free parameters to be fitted.

The model is fitted to each time series voxel in the DCE-MRI data set, using a non-linear optimisation algorithm (Levenberg-Marquardt). For each voxel in turn, the critical point

$$W_p \coloneqq arg\ \max_t f(t) = \left(\frac{b}{c}\right)^{\frac{1}{c}} \tag{5.2}$$

is computed and then normalised as follows

$$\overline{W_p} \coloneqq e^{-\frac{W_p^2}{T^2}}, \tag{5.3}$$

where $T$ is the acquisition duration. The normalisation ensures that $\overline{W_p}$ lies in the interval [0,1). The resulting volume of $\overline{W_p}$ values is filtered, slice-wise, using a 3×3 Gaussian($\mu = 0, \sigma^2 = 1$), to mitigate rapid changes caused by noise and/or patient movement. The resulting matrix of filtered $\overline{W_p}$ values constitutes the CPM. In essence, the CPM measures the speed of contrast uptake at each voxel location, independent of the actual intensity value of the voxel.

Algorithm 5.1: The proposed segmentation algorithm

| | |
|---|---|
| **Inputs** | • A breast DCE-MRI data set comprising one pre-contrast volume and several spatially co-registered post-contrast volumes<br>• Time threshold $\tau$<br>• Relative intensity thresholds $Th_1$ and $Th_2$ |
| **Output** | • A set of connected components locating regions of suspicious enhancement |

**Steps**

1. Compute the 'Critical Points Map' (CPM). This is a map of (normalised and filtered) critical points obtained from a voxel-wise fit of the empiric parametric model of enhancement
2. Compute the 'Contrast Enhancement Image' (CEI). This is a map of the maximum contrast enhancement attained for each voxel within $\tau$ minutes of contrast agent injection
3. Create the 'Domain of Interest' (DOI) binary mask. This mask is derived from the CEI and CPM and defines a subset of the breast volume in which region growing/merging will be performed
4. Identify seed voxels within the DOI whose CEI value is greater than $Th_1$ times its pre-contrast value
5. Perform seeded region growing using the selected seeds, assimilating each voxel neighbour whose CPM value differ by less than $Th_2$ from the average CPM of the region (Algorithm 5.2)
6. Merge abutting regions with similar average CPM (difference smaller than $Th_2$)

### 5.4.2 **Computing the 'Contrast Enhancement Image' (CEI)**

The value of the CEI at voxel position *x* is computed as follows:

$$CEI_x \coloneqq \max\{(I_x^{max} - I_x^0), 0\}, \tag{5.4}$$

Where $I_x^{max} \coloneqq \max_{t\in[0,\tau]}\{I_x^t\}$ is the maximum intensity value at position $x$ during the first $\tau$ minutes of the acquisition, and $I_x^0$ is the intensity value of the voxel at position $x$ in

the pre-contrast image. The assumption is that significant contrast enhancement will occur in a suspicious lesion within $\tau$ minutes after the injection of the contrast agent. A suitable value for $\tau$ is 2–3 minutes as prescribed in the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) MRI lexicon (Morris and Liberman, 2005). The CEI essentially measures the maximum uptake of contrast agent at each voxel during the first $\tau$ minutes of the acquisition.

### 5.4.3 **Computing the 'Domain of Interest' (DOI) binary mask**

The DOI mask, *M,* is obtained by means of entropy-based thresholding (Abutableb, 1989). Two variants of this step were devised to handle DCE-MRI data with and without fat suppression:

1. *Non-fat-suppressed DCE-MRI data.* The DOI is obtained from a two-dimensional (2D) entropy thresholding based on the CEI and the CPM
2. *Fat-suppressed DCE-MRI data.* The DOI is obtained from a one-dimensional (1D) entropy thresholding of the CEI. In our experience, the fat-suppressed data are noisier than the non-fat-suppressed data leading in turn to noisier CPMs and the possibility of under-segmentation if 2D entropy thresholding is used.

The resulting binary mask, *M*, is assumed to be a superset of the set of all suspiciously enhancing regions.

### 5.4.4 **Identifying a seed voxel**

To segment each region, a seeded region growing (Adams and Bischof, 1994) is performed. The method is based on iteratively growing a seed pixel into a region and is described in detail in Appendix B.

At the *i*-th iteration, the candidate seed, $s_i$, is an unlabelled voxel satisfying:

1. $s_i \in M$,
2. $\left[\frac{CEI_{s_i}}{I^0_{s_i}} - 1\right] > Th_1$, and
3. $s_i = arg\max_x(CEI_x)$

where $Th_1$ is a pre-defined threshold defining the minimum relative contrast enhancement rate of a suspicious lesion within τ minutes; $I^0_{s_i}$ is the intensity of the voxel at the position $s_i$ in the pre-contrast image, and $CEI_{s_i}$ is the *CEI* value at voxel position

$s_i$. The value for $Th_1$ is determined experimentally, because the initial enhancement rate is affected by the variation in the pulse sequence, the field strength of the magnet, timing of the contrast agent, the start of the first post contrast scan, and the vendor-related issues (Morris and Liberman, 2005).

### 5.4.5 **Perform seeded region growing**

A seeded region growing (Adams and Bischof, 1994) is performed using a candidate seed, constrained to the mask *M*. For each voxel, *x*, neighbouring the seed region, $R_i$, the similarity criterion is defined by:

$$\delta(x, R_i) \coloneqq |CPM(x) - \overline{CPM}(R_i)| < Th_2, \tag{5.5}$$

where $\overline{CPM}(R_i)$ is the mean critical point value of the evolving region; and $Th_2$ is a predefined similarity threshold, measured experimentally. With the appropriate definition of the neighbour relation, this step can be performed in either 2D or 3D. The algorithm of this step is described in Algorithm 5.2.

Algorithm 5.2: Algorithm for the seeded region growing/merging step

While there are potential seeds:

Find the next potential seed, $s_i$, that satisfies the following criteria:

1. $s_i \in M$
2. $\frac{CEI(s_i)}{I^0_{s_i}} > Th_1$

Push $s_i$ into a priority queue, Q

While Q not empty:

Pop the first $s_i$ from Q

If there is a *j* such that $s_i \in R_j$ then continue (pop next seed)

Mark $s_i$ as part of a new region, $R_i$

Add each pixel $p_j$ that satisfies the following criteria to $R_i$:

1. $p_j \in N(R_i)$
2. $p_j \in M$
3. $p_j \notin R_k, \forall k$
4. $\left|\overline{CPM}(R_i) - CPM(p_j)\right| < Th_2$

### 5.4.6 **Merge abutting regions**

Every pair of neighbouring regions $(R_l, R_m)$ satisfying $|\overline{CPM}(R_l) - \overline{CPM}(R_m)| < Th_2$ are merged into a single region. With the appropriate definition of the neighbour relation, this step can be performed in either 2D or 3D. Moreover, the relation can be defined such that two regions are considered neighbours not only if they abut, but also if they are within a prescribed proximity to one another.

## 5.5 Evaluation of the proposed segmentation algorithm

The aim of the experiment is to evaluate the sensitivity of the proposed segmentation algorithm to enhancing lesions. Two separate sources of DCE-MRI data were used to evaluate the segmentation algorithm: 20 fat-suppressed data sets originating from routine unilateral breast MRI examinations; and 4 non-fat-suppressed data sets from unilateral breast MRI examinations of 4 subjects, each with a diagnosis of primary breast cancer and each imaged prior to intervention. The fat-suppressed data were acquired on a 1.5T Signa EchoSpeed (GE Medical Systems, Milwaukee, USA). The non-fat-suppressed data were acquired on a 1.5T Siemens Sonata (Siemens Medical Systems, Berlin, Germany). All data sets were acquired using a three dimensional FSPGR sequence and an open breast coil that permitted the subject to lie prone. Gadopentate dimeglumine, 0.2 mmol/kg, was administered at a rate of about 3 ml/second. Scan/sequence details are given in Table 5.1. In the case of the fat-suppressed data, the screen captures of suspicious ROIs drawn by the interpreting radiologist, in addition to the MRI findings and histopathology findings (where available), accompanied each data set.

Table 5.1: Scan/sequence parameters for the data sets used to evaluate the segmentation algorithm

| | Fat suppressed data (20 data sets) | Non-fat-suppressed data (4 data sets) |
|---|---|---|
| Acquisition matrix | 256×256 | 256×256 |
| Image matrix | 512×512 (ZIP512) | 256×256 |
| Flip angle | 10° | 25° |
| No. sagittal slices | 24–72 | 48 |
| Slice thickness | 2.5 mm (ZIP2) | 2.5 mm |
| No. postcontrast volumes | 4–10 | 5 |
| Pixel spacing | 0.39 mm | 0.7 mm |
| Labelled ROIs | Yes | No |
| No. labelled ROIs | 1–9 | NA |

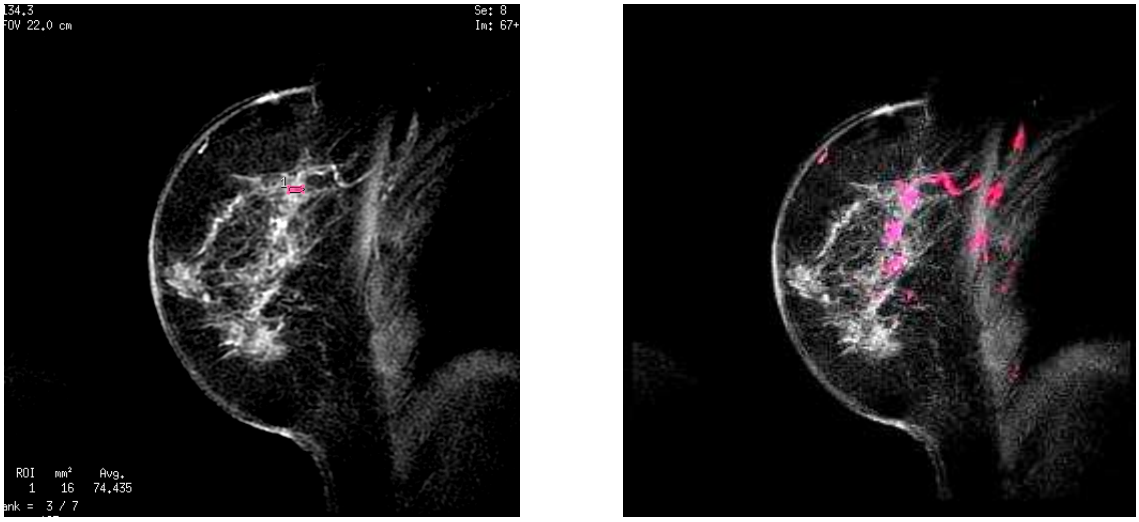
The segmentation algorithm was applied to all 24 data sets using the following parameters, $\tau$=3 minutes, $Th_1 = 0.4$, and $Th_2 = 0.2$. The value for $Th_1$ was based on reported contrast-enhancement rates from the literature, and measurements on additional DCE-MRI datasets originating from routine clinical examination, while the value for $Th_2$ was determined experimentally (using the same datasets as for $Th_1$). The region growing was performed slice by slice in 2D using a 4-connected neighbourhood, and the region-merging step was performed in 3D using a 6-connected neighbourhood to allow consistent distances between the neighbouring voxels. For the 20 subjects for which ROIs had been marked by the interpreting radiologist, the results of the segmentation algorithm included the marked ROIs in all of the cases, i.e. the segmentation algorithm has 100% sensitivity for suspicious ROIs. However, in most cases, additional regions were selected by the algorithm, including the liver (when present in the data). A typical segmentation result is shown in Figure 5.1 and the worst case of over-segmentation is shown in Figure 5.2. The ROIs selected by the interpreting radiologist were not exhaustive and thus may not include all suspicious tissue. In addition, the ROIs delineate the 'hot spots' of enhancement rather than the entire lesions. Consequently, it is not possible to assess the specificity of the segmentation algorithm for the detection of suspicious lesions. Instead, to assess the efficacy of the method, for each labelled data set, the percentage of voxels segmented in each slice was computed for the slices containing a radiologist-marked ROI. For the 20 data sets, the results ranged from 0.20% to 8.98% with a median value of 0.8% and a mean value of 1.9% of the slice image area (Table 5.2). For the four non-fat-suppressed data sets, each corresponding to a subject with diagnosed breast cancer, the segmentation results were visually compared to the respective raw subtraction images. In each case, the segmentation algorithm detected the malignant tissue, whilst at the same time labelling only a small fraction of the voxels in each slice as suspicious (consistent with the findings based on the 20 fat-suppressed data sets).

Table 5.2: Summary of the segmentation results for the 24 DCE-MRI data sets

| Subject | No. of radiologist-marked ROIs * | % of radiologist's ROIs included in the segmentation | % of Image Segmented** | Overall MRI finding | Fat Suppression |
|---|---|---|---|---|---|
| 1 | 1 | 100 | 2.02 | Malignant | Y |
| 2 | 1 | 100 | 2.04 | Malignant | Y |
| 3 | 1 | 100 | 0.20 | Malignant | Y |
| 4 | 7 | 100 | 0.53 | Malignant | Y |
| 5 | 3 | 100 | 0.49 | Malignant | Y |
| 6 | 3 | 100 | 0.99 | Malignant | Y |
| 7 | 1 | 100 | 0.36 | Malignant | Y |
| 8 | 6 | 100 | 0.49 | Malignant | Y |
| 9 | 1 | 100 | 0.52 | Malignant | Y |
| 10 | 3 | 100 | 0.74 | Malignant | Y |
| 11 | 6 | 100 | 2.00 | Benign | Y |
| 12 | 5 | 100 | 8.98 | Benign | Y |
| 13 | 3 | 100 | 0.85 | Benign | Y |
| 14 | 3 | 100 | 1.57 | Benign | Y |
| 15 | 7 | 100 | 5.00 | Benign | Y |
| 16 | 9 | 100 | 6.67 | Benign | Y |
| 17 | 6 | 100 | 0.39 | Benign | Y |
| 18 | 4 | 100 | 0.79 | Benign | Y |
| 19 | 5 | 100 | 0.28 | Benign | Y |
| 20 | 2 | 100 | 4.39 | Benign | Y |
| 21 | - | - | 0.66 | Malignant | N |
| 22 | - | - | 0.26 | Malignant | N |
| 23 | - | - | 4.20 | Malignant | N |
| 24 | - | - | 0.86 | Malignant | N |

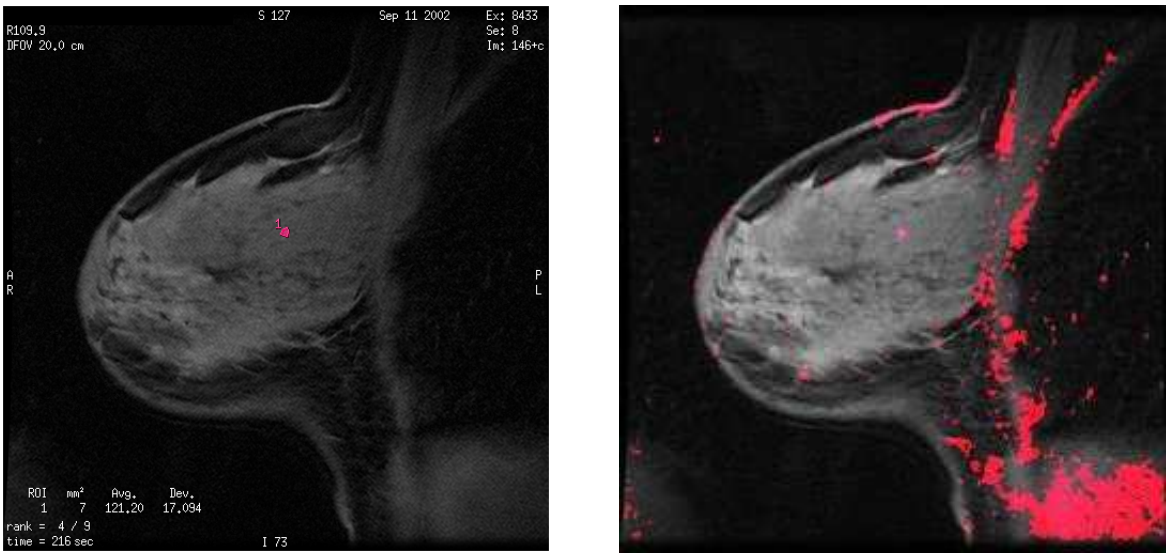
* *The ROIs selected by the radiologist are not exhaustive and thus may not include all of the suspicious lesions*

** *The percentage is calculated for the slices on which the radiologist traced a ROI. The data for subjects 21–24 did not include radiologist-marked ROIs. Consequently, the percentage was calculated based on slices exhibiting a significant amount of segmented voxels.*




(a) (b)

Figure 5.1: Typical segmentation result.
(a) Original slice image superimposed with the ROI drawn by the interpreting radiologist, (b) Segmented regions




(a) (b)

Figure 5.2: Worst case segmentation results.
(a) Original slice image superimposed with the ROI drawn by the interpreting radiologist, (b) Segmented regions. The over-segmentation includes the liver (lower right).

### 5.5.1 **Discussion and conclusion**

The results demonstrate that the proposed segmentation algorithm has 100% sensitivity for the detection of suspicious regions independently identified by a radiologist. The results show that, in the worst case of over-segmentation, the percentage of voxels identified by the algorithm in a given 2D slice is less than 9%. Moreover, this figure is inflated because the algorithm segments the liver in addition to the suspiciously-enhancing tissue. Thus, the algorithm has the potential to be a tool that can assist the clinician with the task of locating suspicious tissue faster and more accurately than can be done by the current manual practice. The increased speed is afforded because the clinician need only review a small sub-volume of the data identified by the algorithm, and the improved accuracy is afforded because the clinician has a much lower chance of missing diagnostically significant regions of suspicious enhancement. In addition, the regions detected by the algorithm can be used as input to a computer-assisted diagnostic system for generating quantitative features for automatic classification of suspicious tissue. Future research includes, (i) the development of a chest wall segmentation algorithm so that enhancement not within the breast volume, e.g. the liver, can be easily distinguished and, (ii) an evaluation of the specificity of the segmentation algorithm based on a manual segmentation by a radiologist of the suspicious tissue in raw DCE-MRI data.

The proposed algorithm is highly sensitive to enhancing lesions. However, the specificity of the algorithm should still be improved. This can be achieved in several ways:

1. Implementing a more fastidious seed selection process
2. Use contextual information, including breast volume segmentation, to filter out lesions that are outside of the breast volume (i.e. out of context), by segmenting the breast volume as a pre-processing step
3. Perform a post-processing step, based on statistical pattern recognition tools to classify and locate those lesions that may have suspicious characteristics and thus may be of interest to the radiologist.

In this thesis, the last option was chosen and hence, a classification method for suspicious lesions has been developed and is presented in Chapter 6.

## 5.6 **Summary**

This chapter reviewed existing methods for the segmentation of suspicious lesions in breast MRI and presented a new, fully-automatic, algorithm for segmenting suspicious lesions in DCE-MRI of the breast. An experimental evaluation of the algorithm's sensitivity to enhancing lesions was also presented. The results of the experiment suggest that the algorithm has a high sensitivity to enhancing lesions in DCE-MRI of the breast. However, it specificity is low and needs to be improved. This can be done by creating a mask of the breast volume as a pre-processing step. This will prevent segmentation of tissue outside of the breast volume, such as the liver that tends to be accidently segmented by the proposed algorithm. Furthermore, as a post-processing step, the segmented tissue can be classified, using statistical pattern recognition techniques. Chapter 6 of the thesis presents new features for automatic classification of suspicious lesions in DCE-MRI of the breast. A classification system, based on these features, may allow filtering the results of a segmentation process and thus improve its specificity.

# 6. Automatic classification of suspicious lesions in DCE-MRI of the breast

The preceding chapters have dealt with low and intermediate-level processing needed to implement a CAE system for breast MRI. This chapter focuses on the classification (high-level processing) of suspicious lesions. It introduces the background and motivation for the automatic classification of the suspicious lesions and reviews the existing methods and features for the classification of the suspicious lesions in breast MRI. It also presents a study whose aim was to determine the most discriminatory subset of features for suspicious lesions in DCE-MRI of the breast. The study shortlists the most discriminatory features for the classification of breast lesions in DCE-MRI that have been reported to date (Section 6.2). It also presents several new features based on the empirical model of contrast enhancement described in Chapter 3, while taking into account the curse of dimensionality, and assessing the classification performance using separate validation data (Section 6.3). The results (presented in Section 6.3.6) suggest that textural and kinetic features are more important than morphometric ones and that the CAE can, indeed, improve the specificity of breast MRI. The core material in this chapter was presented at 'Digital Image Computing: Techniques and Applications', 2009, Melbourne, Australia (Gal et al., 2009a) and was submitted to the Elsevier Artificial Intelligence in Medicine.

## 6.1 Introduction

Dynamic contrast-enhanced (DCE) MRI is being increasingly used in the clinical setting to help detect and characterise tissue, suspicious for malignancy (Sinha et al., 1997). In an attempt to reduce the subjectiveness of the interpretation, the American College of Radiology developed the BI-RADS (Breast Imaging Reporting and Data System) MRI

lexicon (American College of Radiology, 2006) that provides a standard terminology for reporting breast MRI findings (see Appendix D). In particular, the BI-RADS lexicon provides terminology for describing the morphology of a lesion, in addition to its enhancement and kinetic behaviour. A plethora of features for the automatic segmentation of suspicious lesions have been proposed for breast MRI (Sinha et al., 1997, Chen et al., 2006b, Chen et al., 2004, Arbach et al., 2004, Nattkemper et al., 2005, Torheim et al., 2001, Lee et al., 2007). This chapter reviews these and identifies those features that have been found to be the most discriminatory (Section 6.2).

## 6.2 Classification of suspicious lesions in breast MRI

The automatic discrimination between malignant and benign tissue is being researched in several different fields of medical image analysis including: cytology (analysis of cells from a small sample from a suspicious tissue), histology (analysis of cells from an extracted suspicious tissue), MRI, CT, and X-ray mammography. One of the goals of automatic classification is to eliminate the subjectivity that exists in human interpretation by providing quantitative evaluation of features. A compendium of such features developed for automated cytometry can be found in (Rodenacker and Bengtsson, 2003). These are principally features for two-dimensional grey-scale images. Rodenacker and Bengtsson (Rodenacker and Bengtsson, 2003) classify these features into the following four categories, (a) morphometric features, expressing size and shape, (b) densitometric features, expressing total intensity, (c) textural features, expressing a quantification of local intensity variability and, (d) Structural or contextual features, expressing the relationships between different objects. Many of these have been used to date in MRI research. In DCE-MRI, however, contrast enhancement makes the definition of additional enhancement/kinetic features possible. A review of the previous research on breast lesion classification in MRI follows.

### 6.2.1 Features proposed for classification of breast MRI lesions

Various features have been proposed in the literature for the classification of breast lesions in MRI. Many of these quantify identical or similar features that are proposed in the ACR BI-RADS lexicon. These features include:

***Morphometric features***

1. Shape compactness (Sinha et al., 1997, Arbach et al., 2004)
2. Radial entropy, which is measured on the histogram of the normalised radial lengths for all the points on the boundary of the lesion (Sinha et al., 1997)
3. Total bending energy, which is defined in terms of the integral of the curvature at the different points of the contour (Sinha et al., 1997)
4. Ratio of minimum to maximum radial length (Sinha et al., 1997)
5. Maximum standard deviation of the radial gradient histogram (Chen et al., 2004)
6. Circularity (Chen et al., 2004)
7. Irregularity of lesion surface (Chen et al., 2004)
8. Margin gradient along the lesion surface (Chen et al., 2004)
9. Variance of margin gradient along the lesion surface (Chen et al., 2004)
10. The volume of the region (Arbach et al., 2004)
11. Normalised radial length (normalised by the maximum radial length in the lesion) (Arbach et al., 2004)
12. Perimeter length (Arbach et al., 2004)

***Textural features***

13. Textural features based on a co-occurrence matrix (Haralick et al., 1973). These may include contrast, sum of average, sum of variance, sum of entropy, entropy, difference of entropy, inverse difference moment (IDM) and angular second moment (ASM) (Sinha et al., 1997)
14. Level of spiculation (Arbach et al., 2004).

***Kinetic features (based on the temporal data or a fitted model of enhancement)***

15. Total percentage intensity enhancement (Sinha et al., 1997)
16. Slope factor of the enhancement curve (Sinha et al., 1997)
17. Time to half maximum enhancement (Sinha et al., 1997)
18. Maximum rate of enhancement (Sinha et al., 1997)
19. Standard deviation of the means of intensity at each of the time points from the mean intensity of the entire time-intensity curve (Sinha et al., 1997)
20. Minimum enhancement (during the first 30 sec) (Torheim et al., 2001)
21. Wash-in rate (Torheim et al., 2001, Chen et al., 2004, Chen et al., 2006b)

22. Wash-out rate (Torheim et al., 2001, Chen et al., 2004, Chen et al., 2006b)
23. Change in baseline signal (Torheim et al., 2001)
24. Time to peak(Torheim et al., 2001, Chen et al., 2004, Chen et al., 2006b) Maximum uptake (in a pre-defined time frame) (Chen et al., 2004, Chen et al., 2006b)
25. Maximum enhancement-variance (Chen et al., 2004)
26. Enhancement-variance in the first post-contrast volume (Chen et al., 2004)
27. Percentage enhancement (PE) (Lee et al., 2007), on the 3TP curve (Furman-Haran and Degani, 2002)
28. Signal enhancement ratio (SER) (Lee et al., 2007), on the 3TP curve (Furman-Haran and Degani, 2002).

***Other features***

29. Patients' age (Sinha et al., 1997)
30. A variety of features from BI-RADS, where the quantified values are estimated manually by skilled radiologists (Nattkemper et al., 2005). These may include:
    a. Normalised, sub-sampled, version of the enhancement curve
    b. Pattern of enhancement (e.g. No enhancement, centrifugal, heterogeneous, homogenous and ring-like)
    c. Wash-out type (e.g. typical benign, suspicious and typical malignant);
    d. Morphological features: edge (e.g. well defined, poorly defined, spiculated) and contour (e.g. irregular, round).

### 6.2.2 **Review of existing classification models for breast MRI lesions**

The various features can be broadly divided into two classes, static and dynamic. Static features are those that can be derived from a single image volume. They include morphometric (quantifying size, shape, boundary, and so on), intensity (derived from the intensity histogram), and textural (quantifying the spatial variation of intensities). Dynamic features are those derived from the sequence of image volumes acquired before and after contrast agent injection. They include direct measures (e.g. time-to-peak) obtained directly from the temporal data (voxel enhancement curves) and fitted model parameters obtained from empiric-parametric and pharmacokinetic models.

Table 6.1 lists the highest performing classifiers reported in the literature for breast lesion classification in MRI. The performance is measured in terms of either the reported area under the ROC curve (AUC) (Arbach et al., 2004, Chen et al., 2006b, Chen et al., 2004, Lee et al., 2007, Nattkemper et al., 2005) or a single reported sensitivity and specificity (Sinha et al., 1997, Torheim et al., 2001). The table also lists the types of features and types of classifiers used in each case, the number of malignant and the number of benign data samples, the total number of features that were tested and the number of features that were used to generate the best reported result. From the table, it is possible to make the following observations. First, in (Sinha et al., 1997, Arbach et al., 2004) the number of features that were used is arguably too large relative to the number of data samples, in order to avoid the 'curse of dimensionality' (Jain et al., 2000). Second, in the cases where fewer features were used, either the results are not as good as those with more features or the feature values were estimated manually by clinicians rather than automatically (Nattkemper et al., 2005). Thirdly, in (Chen et al., 2006b, Chen et al., 2004, Nattkemper et al., 2005, Sinha et al., 1997, Torheim et al., 2001) it is apparent that the same data was used both for feature selection and for measuring classification performance. This fact may positively bias the results.

Clearly, what is the most useful set of features for breast lesion classification and what performance can be expected using this feature set remain open questions. This question has motivated the present study.

## 6.3 Empirical study of the most discriminatory feature set for lesion classification in breast MRI

### 6.3.1 Methodology

The aim of this study was to determine the most discriminatory subset of features from among those in Table 6.1 and a couple of new features, described in Section 6.3.3, using clinical DCE-MRI data, but taking into account the curse of dimensionality, and assessing the classification performance using separate validation data.

Two experiments were undertaken. The aim of the first experiment was to select the most discriminatory subset of features out of a shortlist of features drawn from Table 6.1 and a couple of new features described below. Only classifiers with a simple

discrimination boundary (i.e. linear in the feature space) were used. The underlying assumption was that a feature set with a high discriminatory power should provide good discrimination between malignant and benign lesions, even when using a simple discrimination boundary. The aim of the second experiment was to estimate the performance of the selected features on unseen data, using the same classifiers from Experiment 1.

### 6.3.2 **DCE-MRI data**

A total of 48 breast lesions from 39 DCE-MRI datasets obtained from clinical breast MRI examinations of 39 different patients were used. The DCE-MRI data were acquired on a 1.5T Signa Echospeed (GE Medical Systems), using a three-dimensional fast-spoiled gradient-echo (3D FSPGR) sequence. The data were acquired using three different protocols because of the protocol changes over time in the clinic from which the data were sourced. The data were divided into two sets.

1. Dataset 1 - 21 lesions from 20 subjects, 11 malignant and 10 benign. The protocol used here involved acquiring a single pre-contrast volume and 5–7 post-contrast volumes, following an injection of a gadolinium-based contrast agent. Images were acquired and viewed at a size of 256x256 pixels per slice, with a pixel size of 0.859 mm, a slice thickness of 1.5 mm and a temporal resolution of 60–90 seconds per volume.
2. Dataset 2 – a total of 27 lesions from 19 subjects. The data were acquired using two different protocols:
    a. 15 lesions from 10 subjects, 8 malignant and 7 benign. The protocol used here involved acquiring a single pre-contrast volume and 4 post-contrast volumes, following an injection of a gadolinium-based contrast agent. Images were acquired at a size of 512x512 pixels per slice, with a pixel size of 0.644 mm, a slice thickness of 1.2 mm, and a temporal resolution of 90–120 seconds per volume
    b. 12 lesions from 9 subjects, 6 malignant and 6 benign. The protocol used here involved acquiring a single pre-contrast volume and 3–4 post-contrast volumes, following an injection of a gadolinium-based contrast agent. Images were acquired at a size of 512x512 pixels per slice, with a pixel size of 0.586-

0.644 mm, a slice thickness of 1.0–1.4 mm and temporal resolution of 90–120 seconds per volume.

All lesion findings (i.e. malignant/benign) were confirmed by cyto- or histo-pathology. Each lesion was manually delineated in the DCE-MR images by a clinical radiographer, using the region growing tool in Osirix (Rosset et al., 2004) and saved as a binary mask.

The data were split into two datasets so that approximately half of the data was used for feature selection and the rest for validation. The data were divided based on the different protocols to strengthen the validation of the selected features. Such a validation increases the chances that the selected features will be useful on a large range of protocols rather than on the protocols that were used for the feature selection experiment or for the classifier training.

Table 6.1: Literature review summary

| No. | Source | Feature types | Feature selection | Classifiers |
|---|---|---|---|---|
| 1 | (Sinha et al., 1997) | Kinetic, morphometric, textural | Y<br>Stepwise | Stepwise linear discriminant analysis |
| 2 | (Torheim et al., 2001) | Kinetic (defined from the mean enhancement curve in a lesion) | Y<br>Exhaustive | Minimum enhancement threshold<br>Fisher's linear discriminant analysis<br>Probabilistic neural network<br>Back propagation neural network<br>Correlation coefficient (CC) |
| 3 | (Chen et al., 2004) | Kinetic, morphometric | Y<br>Stepwise | Linear discriminant analysis |
| 4 | (Arbach et al., 2004) | Morphometric ($T_2$ weighted MRI) | N | Back propagation neural network |
| 5 | (Nattkemper et al., 2005) | Kinetic, morphometric | Y<br>- | K-means<br>Self organising maps<br>Linear discriminant analysis<br>K-nearest neighbour<br>Support vector machine<br>Decision tree |
| 6 | (Chen et al., 2006b) | Kinetic (on a selected kinetic curve) | Y<br>Single*** | Fuzzy C-mean |
| 7 | (Lee et al., 2007) | Kinetic (based on 3TP curve (Degani et al., 1997)), classify voxel-wise | N | K-means |

| No. | No. data samples | No. Benign | No. Malignant | Validation method | Test set size | SE* | SP* | Best AUC | **Total features |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 43 | 23 | 20 | N/A | N/A | 95% | 93% | N/A | 10/18 |
| 2 | 127 | 57 | 70 | Leave-one-out cross-validation | Leave-one-out | 81% | 95% | N/A | 2/5 |
| 3 | 121 | 44 | 77 | Leave-one-out cross-validation | Leave-one-out | | | 0.86±0.04 | 4/14 |
| 4 | 35 | N/A | N/A | 40% - holdout | 40% - randomly selected | | | 0.91±0.05 | 5/5 |
| 5 | 74 | 25 | 49 | Leave-one-out cross-validation | Leave-one-out | | | 0.88±N/A | *§2/6* |
| 6 | 121 | 44 | 77 | N/A | N/A | | | 0.85±0.04 | 1/4 |
| 7 | 25 | - | 25 | Visual only – classify voxels | N/A | Visual only – classify voxels | | | 2/2 |

* *SE = Sensitivity, SP = Specificity*

** *Number of features used to achieve the best reported result / total number of features tested*

*** *In this case every single feature out of four was tested separately*

§ *Feature values were estimated manually, by experienced radiologists*

### 6.3.3 **Features considered**

In practice, clinicians commonly analyse each volume slice-by-slice (i.e. in 2D). Thus, it is reasonable to assume that the information in the 2D slices is sufficient for discriminating between benign and malignant. Given that the pixels in the data that were used for this study are not cubic and, to avoid interpolating the original data, features were extracted in 2D from the slice containing the largest cross section between the binary mask and the original MRI image of each lesion. This dimensional constraint eliminated the use of some morphological features, especially those that account for 3D connectivity.

To keep the set of candidate features for Experiment 1 to a reasonable size, 10 features were shortlisted from Table 6.1. They include examples of morphological (based on the binary mask of the lesion), textural (based on intensities inside the lesion) and kinetic (based on individual/mean enhancement curves inside the lesion) features. In the following feature list, $T_0$ denotes the pre-contrast volume and *MI* denotes an image that is constructed from the maximum intensity projection along the temporal dimension (more specifically, the first 5 minutes of the acquisition). In the general case, $T_0$ may be substituted with any post-contrast volume. Moreover, the maximum intensity for *MI* can be extracted either from the original data or calculated from a model of enhancement.

The selected features are as follows.

***Features shortlisted from the literature***

1. *Eccentricity.* The eccentricity of the ellipse that has the same second moments as the lesion. This is the ratio of the distance between the foci of the ellipse and its major axis length.
2. *Solidity.* The proportion between the number of the pixels that are in the region and the number of pixels in the convex hull of the region (i.e. *Area* divided by *Convex Area*).
3. *Shape compactness.* Defined as (Sinha et al., 1997):

$$c = \frac{p^2}{a}$$

   where $p$ is the perimeter of the lesion and $a$ is the area of the lesion.
4. *Fractal Dimension* (Landini and Rippin, 1996). Mean value of the fractal dimension of the boundary defined as:

$$\alpha = \frac{\log\big(M(\varepsilon)\big)}{\log(\varepsilon)}$$

where $M(\varepsilon)$ is the number of locally connected pixels (using 8-neighbourhood connectivity) in a window of size $\varepsilon$ by $\varepsilon$ pixels.

5. *Fractal Dimension Entropy.* Entropy of the fractal dimension of the boundary.
6. $T_0$ *Entropy.* Entropy of the intensities in the pre-contrast image inside the lesion.
7. *Time-to-Peak Entropy.* Entropy of the time-to-peak ($T_p$) values.
8. *Mean Time-To-Peak.* Time-to-Peak of the mean enhancement curve.
9. *Mean Wash-in rate.* The wash-in slope of the mean enhancement curve (line 1 in Figure 6.1). The wash-in slope is given by $f(T_p)/T_p$, where $T_p$ is the intensity peak time as given by the enhancement curve $f$ and $f(T_p)$ is the peak value (i.e. value of maximum enhancement).
10. *Mean Wash-out rate.* The wash-out slope of the mean enhancement curve (line 2 in Figure 6.1). The wash-out slope is given by $\big(f(T) - f\big(T_p\big)\big)/\big(T - T_p\big)$.

***New features***

11. *MI Entropy.* Entropy of the intensities in the *MI* image inside the lesion.
12. *Gradient Correlation (GC) in* $T_0$ - The mean correlation between adjacent gradients in $T_0$ (Levin et al., 2003):

$$GC_{T_0} = Mean\left(\frac{\langle G_i, G_j\rangle}{\|G_i\| \cdot \|G_j\|}\right)$$

where $G_i = \nabla T_0(x_i, y_i)$, $G_j = \nabla T_0\big(x_j, y_j\big)$, are gradients in the pre-contrast image, $i$ and $j$ are two pixel positions in the image that obey the N4 neighbourhood rule. The gradient values, $\nabla T_0$, were calculated based on the discrete differences over the spatial axes in the N4 neighbours.

13. *Gradient Correlation in MI.* The mean correlation between adjacent gradients in *MI*:

$$GC_{MI} = Mean\left(\frac{\langle G_i, G_j\rangle}{\|G_i\| \cdot \|G_j\|}\right)$$

where $G_i$ and $G_j$ are as above.

***Other features:***

14. *Random number.* A randomly-generated number from uniform distribution, to be used as a mean of validation in the feature selection experiment.

Features 1 to 5 were chosen because they provide a description of the shape and boundary irregularity of the lesion. More specifically, features 1, 2 and 3 can be related to shape descriptions (i.e. round/oval/lobular) in the ACR BIRADS lexicon (American College of Radiology, 2006). Feature 4 can be related to the level of irregularity of the shape or the boundary. The smoother the boundary is, the less irregular it is. However, when only part of the boundary is irregular, the measurement of the fractal dimension alone may yield low values (i.e. present the boundary as regular). Feature 5 compensates for this weakness of feature 4. In the case where only part of the lesion's boundary is irregular, feature 5 will yield a high value. In terms of intensity variation, features 6 and 12 provide two different measures of intensity irregularity in the pre-contrast image. Note, that features 12 and 13 are inspired by an in-painting related work (Levin et al., 2003).

Features 7, 8, 9 and 10 measure properties of the enhancement pattern in DCE-MRI and are commonly used to describe the characteristics of kinetic curves (Nattkemper et al., 2005, Sinha et al., 1997, Lee et al., 2007, Chen et al., 2004). The importance of these characteristics is also indicated in the ACR BIRADS lexicon (American College of Radiology, 2006).

Feature 14 (random number) was added to the initial feature set in order to validate the feature selection process (i.e. a 'sanity check'), assuming that a selection of feature 14 by the process will imply that the results of the feature selection process are unreliable or suggesting features that do not contain real discriminatory information.

### 6.3.4 **Experiment 1 - Determining the best feature subset**

The aim of this experiment was to find the feature subset that produces the best discrimination between malignant and benign tissues while minimising the number of features used. The curse of dimensionality (Jain et al., 2000) motivates the reduction of the number of features used for classification to increase the generalisability of the classification system. The small number of datasets (21 lesions) that was available for

the experiments dictated a feature subset of size no larger than 2. The selected features and classifiers from Experiment 1 were then used in Experiment 2

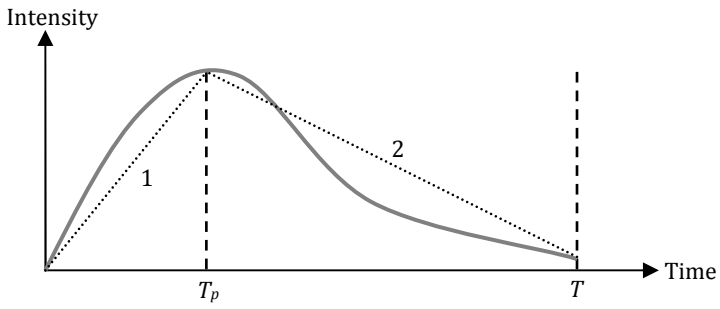



Figure 6.1: The mean wash-in rate (1) and mean wash-out rate (2) of the mean enhancement curve
$\boldsymbol{T_p}$ represents intensity peak time as calculated by the enhancement model. $T$ is the total acquisition time

Before the feature extraction, all clinical data was denoised using the DNLM algorithm that was presented in Chapter 4 of this thesis. For the feature selection, the lesions from dataset 1 were used. Although the relatively small number of training samples dictates that only one, or perhaps two, features can be used for classification, feature sets of size 3 and 4 were also considered in order to assess the consistency of the selection of certain features, rather than to assess their performance.

In order to extract features 11 and 13, the *MI* image was computed. In this study, features 7, 8, 9 and 10 as well as the *MI* image were calculated, based on the empirical model that was presented in Chapter 3 of this thesis and given by:

$$f(t) = a \cdot t \cdot e^{-\frac{t^c}{b}}$$

where $t$ is the time that has elapsed since the injection of the contrast agent; and $a$, $b$ and $c$ are the free parameters. The proposed model has been chosen because it has shown to fit better to clinical MRI data (Gal et al., 2007b) and has a parsimonious form relative to the pharmacokinetic models. The model of enhancement was fitted voxel-wise and was also fitted to the mean enhancement curve of the delineated lesion using the Levenberg-Marquardt non-linear optimisation method (Gal et al., 2007b). Then, the maximum value of the fitted model was calculated for each voxel to create the *MI*. The maximum value of the model is given by:

$$f_{max} = a\left(\frac{b}{c \cdot e}\right)^{\frac{1}{c}}$$

which is the value of the point where $f'(t) = 0$. The point in time where the model gets its maximum value is the time-to-peak point and is given by:

$$T_p = \left(\frac{b}{c}\right)^{\frac{1}{c}}$$

To select the best features, an exhaustive search was used, using several different classifiers. An exhaustive search was selected because it considers all the possible feature subsets and thus finds the global optimum of the feature selection problem (rather than heuristic methods, such as simulated annealing and genetic algorithms). This was possible because of the small number of features (13 features) and the small size of the subsets to be considered (up to four features). In total, for each classifier type that was tested, a total of $C_1^{13} + C_2^{13} + C_3^{13} + C_4^{13} = 1092$ subsets of features were tested.

The feature selection was performed using several different classifiers (Webb, 2003):

1. Logistic regression
2. Fisher's linear discriminant analysis
3. Linear Bayesian discriminant analysis
4. Support Vector Machine (SVM) with the Sigmoid kernel (order 1 and 2).

These were chosen mainly because they are simple to implement and train (e.g. easily converge and do not have a large degree of freedom) and are well studied. All classifiers examined in this experiment create a linear or simple discrimination boundary. The underlying assumption in the selection of classifiers was that an effective selection of the feature set would lead to an easily-separable classification problem and thus would reduce the influence of the bias/variance dilemma (Geman et al., 1992).

In the first stage of the experiment, a set of all the possible subsets of size one to four was generated from the original feature set. Next, each of the classifiers was trained and tested on each of the feature subsets, using a leave-one-out strategy with all 20 training samples. Finally, the AUC and the minimum error of the empirical ROC curve were calculated and recorded for each combination of feature subset and classifier. In the cases where many feature subsets gave the same optimal AUC (usually AUC of 1.0), a frequency table was constructed, showing the occurrence of each feature in the list of optimal feature subsets.

### 6.3.5 **Experiment 2 – Classification validation**

The aim of this experiment was to, (1) validate the classification performance of the best feature sets identified in experiment 1 using unseen data and, (2) validate the classification performance of a variety of classifier types using the best feature sets.

The selected classifiers were trained on dataset 1 and were then tested on dataset 2. The performance of the classifiers was measured in terms of the AUC.

### 6.3.6 **Results**

#### ***Experiment 1 - Feature selection***

The results of the feature selection experiment show that the Fisher's linear classifier (Table 6.5) has the lowest classification performance (in terms of the AUC) while the SVM with sigmoid kernel has the highest. SVM with other kernels (e.g. distance, exponential and radial) did not perform as well as the sigmoid kernel and thus, they are not discussed here. The logistic regression results (see table 6.2) show that the larger the feature set is, the better the classification performance is, while the best results were obtained by using features 7, 8, 12 and 13 from the feature list. The AUC standard error was calculated according to (Hanley and McNeil, 1982), based on the Wilcoxon test. Note that the high AUC standard errors reflect both the small size of the training data set and the inherently large variance of the AUC estimate obtained using the leave-one-out cross validation.

Out of the SVM results, the ones with the sigmoid kernel yielded the highest AUC values when using a sigmoid kernel of order 1 or 2. The most dominant feature in this case is feature 13 (i.e. gradient correlation in *MI*), which, in the case of a single feature based classification, yields the highest AUC value (AUC of $0.963 \pm 0.043$ with min error 0.095 for Sigmoid kernel of order 1 and AUC of 1.0 for Sigmoid kernel of order 2). When testing feature subsets of size 2 or higher, the best subsets yield an AUC value of 1.0. Given the large number of optimal subsets for subsets of sizes 3 and 4, the SVM results (sigmoid kernel) are presented in Table 6.3 and Table 6.4 in terms of the relative frequency for each feature, rather than the list of best subsets, i.e. the number of times that a feature appeared in an optimal feature subset out of the total number of optimal subsets. The tables show only the features that were included in at least one feature set.

Table 6.2: Feature selection results for the logistic regression classifier

The best AUC (in a leave-one-out cross validation) is presented with its calculated error rate. 'Min error' represents the minimum ratio between misclassification and number of test cases among all possible thresholds.

| No. of features | Best AUC ± standard error | Corresponding feature subset | Min error |
|---|---|---|---|
| 1 | 0.654±0.121 | {8} | 0.238 |
| 2 | 0.727±0.111 | {13, 8} | 0.286 |
| 3 | 0.736±0.110 | {13,7, 8} | 0.238 |
| 4 | 0.800±0.098 | {12, 13, 7, 8} | 0.143 |

Table 6.3: Relative frequencies of features in the feature selection experiment for SVM with sigmoid kernel of order 1

Relative frequency 0 represents a feature that was never selected for an optimal subset while relative frequency of 1 represents a feature that was selected in each one of the optimal subsets.

| Feature | Relative frequency | | | |
|---|---|---|---|---|
| | 1 feature | 2 features | 3 features | 4 features |
| 4 | 0.00 | 0.00 | 0.22 | 0.41 |
| 5 | 0.00 | 0.00 | 0.00 | 0.06 |
| 6 | 0.00 | 0.00 | 0.22 | 0.41 |
| 8 | 0.00 | 0.00 | 0.22 | 0.41 |
| 9 | 0.00 | 0.50 | 0.56 | 0.65 |
| 10 | 0.00 | 0.00 | 0.22 | 0.41 |
| 11 | 0.00 | 0.50 | 0.56 | 0.65 |
| 13 | 1.00 | 1.00 | 1.00 | 1.00 |

Table 6.4: Relative frequencies of features in the feature selection experiment for SVM with sigmoid kernel of order 2

| Feature | Relative frequency | | | |
|---|---|---|---|---|
| | 1 feature | 2 features | 3 features | 4 features |
| 4 | 0.00 | 0.17 | 0.36 | 0.41 |
| 5 | 0.00 | 0.00 | 0.14 | 0.31 |
| 6 | 0.00 | 0.17 | 0.36 | 0.41 |
| 8 | 0.00 | 0.17 | 0.36 | 0.41 |
| 9 | 0.00 | 0.17 | 0.43 | 0.52 |
| 10 | 0.00 | 0.17 | 0.36 | 0.41 |
| 11 | 0.00 | 0.17 | 0.43 | 0.52 |
| 13 | 1.00 | 1.00 | 1.00 | 1.00 |

The results for Fisher's linear discriminant are presented in Table 6.5 and show similar performance to logistic regression, with the same optimal feature subsets. Again, as in the case of logistic regression, the high AUC standard errors reflect both the small size of the training data set and the inherently large variance of the AUC estimate obtained using leave-one-out cross validation.

Table 6.5: Feature selection results for the Fisher's linear classifier
The AUC is presented with its calculated error rate. 'Min error' represents the minimum ratio between misclassification and number of test cases among all possible thresholds.

| No. of features | Best AUC ± standard error | Corresponding feature subset | Min error |
|---|---|---|---|
| 1 | 0.636±0.122 | {8} | 0.286 |
| 2 | 0.664±0.120 | {13,8} | 0.286 |
| 3 | 0.773±0.104 | {13,7,8} | 0.190 |
| 4 | 0.836±0.090 | {12,13,7,8} | 0.142 |

The results of the linear discriminant Bayes normal classifier (LDC), in Table 6.6, show that feature 13 (correlation between adjacent gradients in the maximum enhancement image) is the best feature for a single feature classification. However, the performance of the classifier decreases as the size of the feature space increases. Also, the minimum error rate (estimate of the probability of misclassification) increases as the size of the feature space increases. This illustrates the peaking phenomenon or curse of dimensionality (Jain et al., 2000), wherein the performance of a classifier increases to a peak as the number of features is increased, but then decreases.

Table 6.6: Feature selection results for the LDC classifier
The AUC is presented with its calculated error rate. 'Min error' represents the minimum ratio between misclassification and number of test cases among all possible thresholds.

| No. of features | Best AUC ± standard error | Corresponding feature subset | Min error |
|---|---|---|---|
| 1 | 1.000±0.000 | {13} | 0.000 |
| 2 | 0.818±0.094 | {12,13} | 0.190 |
| 3 | 0.745±0.109 | {11,12,7} | 0.238 |
| 4 | 0.770±0.104 | {1,5,11,12} | 0.286 |

### *Experiment 2 – Classification validation*

The classification performance was measured on the same classifiers that were used for Experiment 1. The best features from Experiment 1 were selected to create the feature spaces for Experiment 2. Feature 13, which demonstrated the best result, was validated as a single feature in addition to each of the features 11, 8 and 9. The results are presented in Table 6.8. The results show that the best performance is given by the logistic regression and the SVM (with Sigmoid kernel) classifiers, while using features 13 and 9. The LDC classifier, on the other hand, shows a very poor generalisability to unseen data. The scatter plot of features 13 and 9 is shown in Figure 6.2. The ROC curve with the highest AUC (of the logistic regression classifier) is shown in Figure 6.3. The

different operating points, for the Logistic Regression, between 100% sensitivity and 100% specificity are presented in Table 6.7.

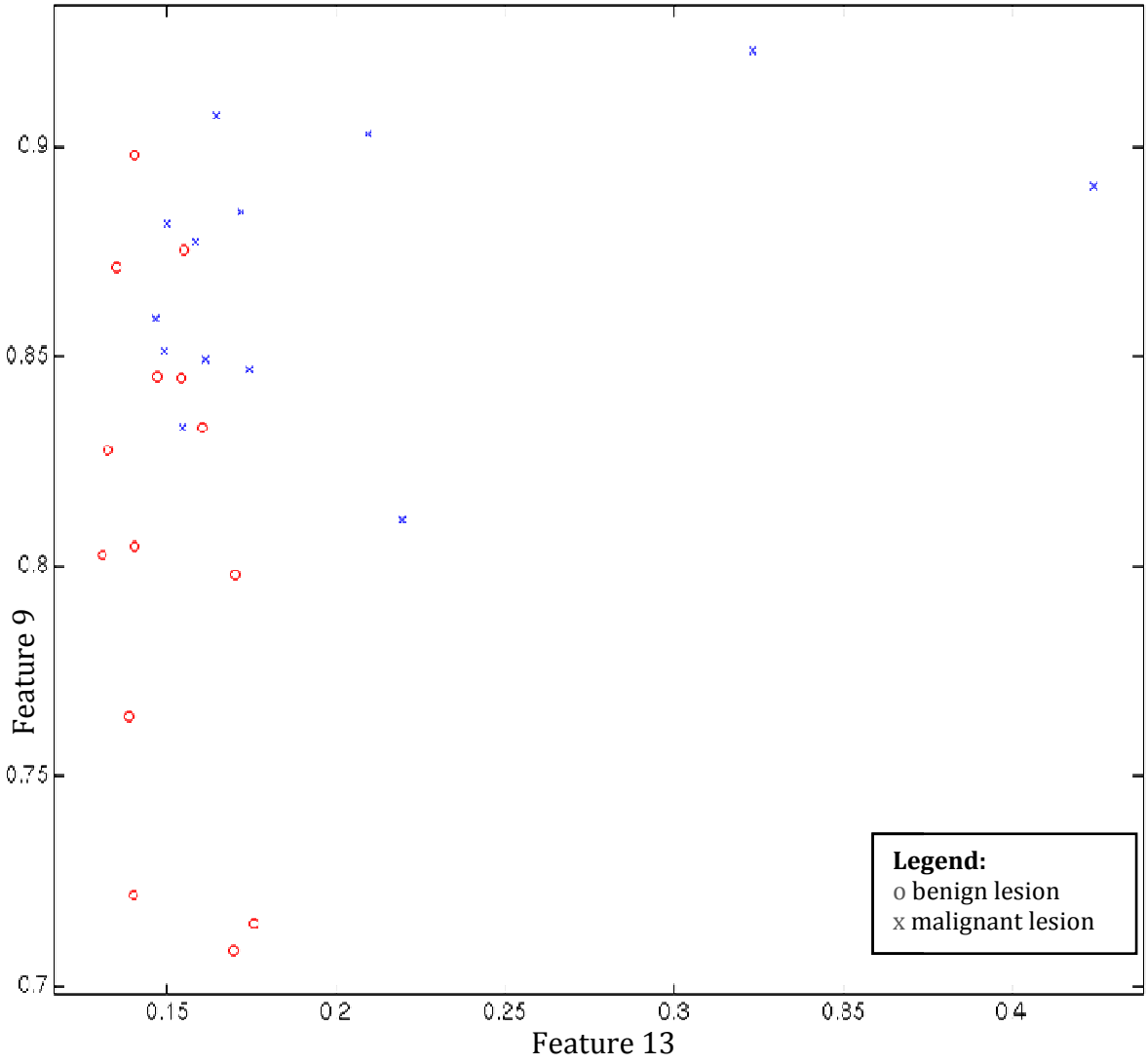



Figure 6.2: Scatter plot of the validation data using features 13 and 9

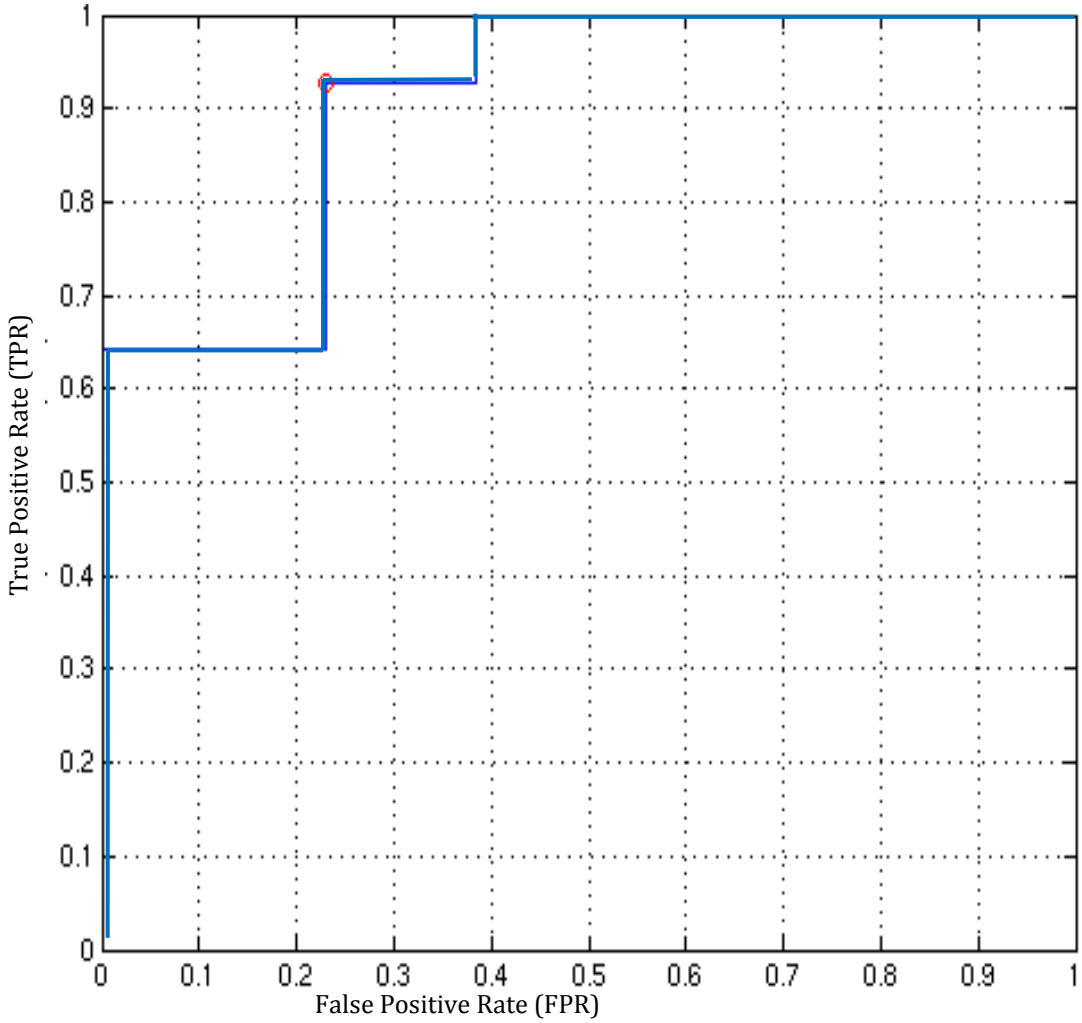



Figure 6.3: The ROC curve of the logistic regression when tested using features 13 and 9. The steps in the curve are related to the small number of data samples in the validation (i.e. test) set.

Table 6.7: Operating points of interest on the ROC curve of the logistic regression classifier, using features 13 and 9

| Threshold | False Positive | False Negative | Sensitivity | Specificity | Positive predictive value (PPV) |
|---|---|---|---|---|---|
| 0.63 | 14 | 8 | 1.00 | 0.62 | 0.74 |
| 0.65 | 13 | 9 | 0.93 | 0.69 | 0.76 |
| 0.66 | 13 | 10 | 0.93 | 0.69 | 0.76 |
| 0.68 | 13 | 10 | 0.93 | 0.77 | 0.81 |
| 0.69 | 12 | 10 | 0.86 | 0.77 | 0.80 |
| 0.70 | 11 | 10 | 0.79 | 0.77 | 0.79 |
| 0.70 | 10 | 10 | 0.71 | 0.77 | 0.77 |
| 0.72 | 9 | 10 | 0.64 | 0.77 | 0.75 |
| 0.73 | 9 | 11 | 0.64 | 0.85 | 0.82 |
| 0.73 | 9 | 12 | 0.64 | 0.92 | 0.90 |
| 0.75 | 9 | 13 | 0.64 | 1.00 | 1.00 |

Table 6.8: Results of classification performance evaluation

| Classifier | Feature subset | AUC | Min error |
|---|---|---|---|
| SVM (Sigmoid 1 kernel) | {13} | 0.80±0.09 | 0.26 |
| SVM (Sigmoid 1 kernel) | {11,13} | 0.88±0.07 | 0.15 |
| SVM (Sigmoid 1 kernel) | {13,8} | 0.80±0.09 | 0.26 |
| SVM (Sigmoid 1 kernel) | {13,9} | 0.90±0.06 | 0.15 |
| SVM (Sigmoid 2 kernel) | {13} | 0.80±0.09 | 0.26 |
| SVM (Sigmoid 2 kernel) | {11,13} | 0.88±0.07 | 0.15 |
| SVM (Sigmoid 2 kernel) | {13,8} | 0.80±0.09 | 0.26 |
| SVM (Sigmoid 2 kernel) | {13,9} | 0.86±0.07 | 0.18 |
| Fisher's LD | {13} | 0.80±0.09 | 0.26 |
| Fisher's LD | {11,13} | 0.86±0.07 | 0.22 |
| Fisher's LD | {13,8} | 0.65±0.11 | 0.26 |
| Fisher's LD | {13,9} | 0.85±0.07 | 0.18 |
| LDC | {13} | 0.55±0.11 | 0.33 |
| LDC | {11,13} | 0.47±0.11 | 0.37 |
| LDC | {13,8} | 0.62±0.11 | 0.33 |
| LDC | {13,9} | 0.21±0.09 | 0.52 |
| Logistic Regression | {13} | 0.80±0.09 | 0.26 |
| Logistic Regression | {11,13} | 0.81±0.08 | 0.22 |
| Logistic Regression | {13,8} | 0.73±0.10 | 0.22 |
| **Logistic Regression** | **{13,9}** | **0.91±0.06** | **0.15** |

### 6.3.7 **Discussion of empirical study results**

The results of Experiment 1 (feature selection), suggest that the correlation between adjacent gradients in the maximum enhancement image, *MI*, (feature 13) is the most important single feature for classifying suspicious lesions in DCE-MRI of the breast. This is, perhaps, not surprising, given that heterogeneous enhancement is a characteristic of malignant lesions (Morris and Liberman, 2005) and feature 13 is a measure of both texture and enhancement. To the best of our knowledge, this is the first time that this feature, which was inspired by image in-painting (Levin et al., 2003), has been used for classifying lesions in breast MRI.

Three other dominant features are the entropy of intensities in the maximum intensity image (feature 11), mean time-to-peak (feature 8) and the mean wash-in rate (feature 9). Given that features 8 and 9 represent similar information, it is not surprising that both turn out to be good features. Note that although features 11 and 13 are textural features, they implicitly rely on kinetic information in the DCE-MRI sequence (i.e. the maximum enhancement intensity value in the first 5 minutes).

Interestingly, the results show that morphometric features tend to be less important. Two possible reasons for this are, (i) the relatively low resolution of DCE MR images which limits the accuracy of boundary description and, (ii) the relative fuzziness of the boundary definition of malignant lesions in MR images (Behrens et al., 2007) that may lead to a subjective manual delineation.

The results of Experiment 2 show that features 13 and 9 constitute the best feature set of size 2, when used with the logistic regression classifier. Hence, logistic regression shows the best generalisation for unseen data when used with features 13 and 9, yielding an AUC value of $0.91 \pm 0.06$ on an unseen validation data. The SVM classifier with the Sigmoid kernel of order 2, shows similar results to the Logistic regression, suggesting that features 13 and 9 indeed have a high discriminatory power for DCE-MRI breast lesions.

## 6.4 Summary and conclusions

This chapter of the thesis reviewed the best performing features and classifiers that have been proposed for classifying benign and malignant lesions in DCE-MRI of the breast. It has been found that textural, kinetic and morphometric features are the most useful types of features for the classification of lesions in DCE-MRI. A set of 13 features were proposed. These are based on textural, kinetic or morphological information, in addition to the features that rely on both textural and kinetic information from the DCE-MRI data. The results of the two experiments were also presented, (i) the feature selection experiment and, (ii) the classification performance evaluation.

The results of the two experiments suggest that suspicious lesions in DCE-MRI of the breast can be classified, with a high AUC of $0.91 \pm 0.06$, using a feature set of two features only. These features include both textural and kinetic information from the lesion. Moreover, morphological features seem not to play an important role in automatic classification of suspicious lesions in DCE-MRI of the breast. The results of the experiments also suggest that combining different types of information (e.g. kinetic and textural) into a single feature may enhance the discriminatory power of a feature space without increasing its size.

# 7. Summary and conclusions

This section reviews the thesis, summarises its key contributions and discusses their implications. It also presents the limitations of the research and opportunities for further research.

## Thesis review

**Chapter 1** described the field of breast MRI and the inherent problems in interpreting breast MRI data. Although MRI is a promising imaging modality for breast cancer, it suffers from poor specificity and subjectivity in interpretation. The chapter also described the different approaches to overcome these weaknesses including, standardisation of reporting terminology (e.g. the ACR BI-RADS lexicon) and computerised analysis.

**Chapter 2** introduced the terminology and background concepts of MR Imaging of the breast and of breast cancer. It also described the basic physics behind the different types of MR imaging in addition to the physiology of the female breast. The chapter also provides an overview of the different types of breast cancer and how they appear in MR images.

**Chapter 3** dealt with the parameterisation of contrast enhancement. It reviewed existing models for contrast enhancement in DCE-MRI and presented a new, empirical, model of enhancement that is parsimonious in form. The chapter also described the properties of the new model and its advantages, in comparison with the existing pharmacokinetic models. A quantitative comparison between the new model and existing models in terms of 'goodness of fit' to measured data and 'goodness-of-fit' to missing temporal data (i.e. prediction of missing points) is also presented. The existing models that were chosen for the comparison were the Tofts, Brix and Hayton models. The results of the comparison showed that, in terms of goodness-of-fit, the proposed

model performs better than the Tofts and Brix models and about equally as well with the more complicated Hayton model.

**Chapter 4** dealt with noise reduction in MRI. It described the nature of noise in MR images and ways to reduce it. The chapter reviewed denoising methods for MRI and described a new denoising method, called Dynamic Non-Local Means (DNLM), designed to exploit the redundancy of information existing in DCE-MRI. The DNLM algorithm was compared to seven other denoising methods quantitatively (in terms of the MSE) and qualitatively (i.e. visually) and the results showed that DNLM performs significantly better than all the competing algorithms. The chapter also presented an experiment that demonstrates a case where the original NL-Means algorithm can actually eliminate small details that may be diagnostically important.

**Chapter 5** dealt with the segmentation of the enhancing tissue in DCE-MRI. It presented a method for automatically segmenting enhancing lesions in DCE-MRI of the breast. The method is based on seeded region growing where the selection of seeds is performed automatically. The algorithm was applied to a set of 24 DCE-MRI datasets from clinical practice with suspicious lesions that were marked by a radiologist. The results showed that the sensitivity of the algorithm to enhancing lesions is 100% on the test datasets. Nevertheless, it is possible to improve the performance of the segmentation, in terms data reduction (i.e. reduce the volume of the breast that is being segmented by the algorithm), using a lesion post-classification, similar to the one proposed in Chapter 6.

**Chapter 6** dealt with the classification of suspicious lesions in DCE-MRI of the breast. It reviewed existing methods for the automatic classification of suspicious lesions in DCE-MRI of the breast. The chapter also described the process of the selection of a best combination of a feature set for classifying suspicious lesions in DCE-MRI of the breast. An evaluation of the performance of the selected features was also presented. Two experiments were presented: (i) feature selection; and, (ii) classification validation. Both experiments were performed on DCE-MRI data from clinical practice and all features were extracted from lesions that were manually delineated by clinical experts. The results show that, based on two features, both derived from the kinetic model of enhancement that was proposed in Chapter 3, a classification system for suspicious lesions can be built, yielding an AUC of 0.9 on the given validation dataset.

## Key contributions and findings

- A new empiric model of contrast enhancement. The new model is parsimonious in form, containing only one exponential factor. It has three free parameters and makes no assumptions about the perfusion of contrast agent between the tissue compartments, as in the case of the pharmacokinetic models.
- A new method to compare the models of contrast enhancement was proposed. The new method suggests that, in addition to traditional measures, such as the MSE to the fitted data, each competing model should be also fitted to a high temporal resolution data where a few data points have been eliminated from the data. The fitted model should then be measured against the missing points to indicate if its 'goodness-of-fit' is related only to flexibility or to its actual ability to represent the truth (i.e. ability to predict missing data points).
- A new denoising algorithm, called DNLM, designed specifically for DCE-MRI. The DNLM algorithm was tested on DCE-MRI data of the breast, but can be used on other types of DCE-MRI datasets. Also, the algorithm has the potential for use in other imaging modalities that produce contrast enhancement, between different acquired volumes, as part of the acquisition process (e.g. nuclear medicine modalities).
- Empirical evidence that medical experts and image processing experts evaluate the quality of denoising algorithms in a similar manner.
- Demonstration that although the NL-Means denoising algorithm is widely used for denoising medical images, it can actually eliminate small details in the image that may be diagnostically important.
- A novel algorithm for segmenting enhancing lesions in DCE-MRI of the breast. The algorithm is fully automatic and is based on the well-known seeded region growing method. The algorithm is sensitive to enhancing lesions and has the potential to assist a radiologist in locating suspicious lesions in DCE-MRI by reducing the volume of the data that need to be scanned.
- Two novel features for the classification of suspicious lesions in DCE-MRI of the breast. Both features are based on the empiric model of enhancement that was proposed in Chapter 3 of the thesis.

- Empiric evidence that morphometric features have little contribution in the classification of benign and malignant tissues in DCE-MRI of the breast. This result is possibly caused by the relatively low resolution of DCE-MR images that do not allow a fine description of the lesion boundary and because of the inherited fuzziness of the lesion boundaries in these types of images.

# Implications of findings

### Denoising DCE-MRI data with DNLM

The empirical results of Chapter 4 suggest that the DNLM has the highest potential for suppressing noise in DCE-MRI images without diminishing clinically important details. An increase in the SNR in the DCE-MR images provides the opportunity to acquire images with a higher spatial or a higher temporal resolution. A higher spatial resolution may reveal more anatomical details, while a higher temporal resolution will provide more accurate kinetic features and will allow fitting models with more free parameters to the data that potentially describe more accurately the perfusion of the contrast agent in the body.

### Automatic classification of suspicious lesions in DCE-MRI

The classification system, proposed in Chapter 6 of the thesis, showed a higher performance in classifying suspicious lesions in DCE-MRI of the breast than the existing classification systems. More specifically, the AUC of the proposed system is $0.91 \pm 0.06$, based on 27 unseen datasets. The operating points on the ROC curve of the proposed classification system suggest that both the specificity and the sensitivity of DCE-MRI of the breast can be improved by using computerised analysis methods (e.g. 93% sensitivity with 77% specificity).

# Limitations of the proposed methods

### Denoising algorithm

The DNLM has three parameters that need to be tuned prior to the denoising, (i) the similarity window size, (ii) the searching distance and, (iii) the weights decay rate ($h$). The first two can be tuned once for each protocol. However, an estimation of the noise in the image is required to determine the value of the parameter $h$. The accurate

estimation of noise in images is influenced by image manipulation, including rescaling and different types of filtering that change the characteristics of the noise in the image. Thus, the noise estimation process, in addition to the denoising process, must receive the image before it has been manipulated in any way that may change the characteristics of the noise.

**Segmentation algorithm**

The segmentation algorithm in Chapter 5 incorporates two different thresholds that need to be specified in advance. The first, $Th_1$, is used for the selection of the seeds. This threshold can be determined, based on previous clinical studies, because it relies on the expected relative enhancement in the intensity of malignant tissue in DCE-MRI. However, the second threshold, $Th_2$, which determines the seed-growing criterion, should be selected empirically. In certain cases, it may be hard to pick a sufficiently general value for this threshold, such that it will suit a large variety of studies and protocols.

Another limitation of the algorithm is the binarization step. This step relies on a global threshold. In cases where there are several enhancing tissues in the image, stronger enhancing tissues may deflect the automatically-selected threshold upwards and so eliminate weaker enhancement from the binary mask. An especially problematic case is when the liver or the heart, which tend to enhance greatly, appears in the image and masks malignant tissue that enhances less.

**Classification of suspicious lesions**

The performance of the classification system in Chapter 6 is higher than those that can be found in the literature, in terms of the AUC. Nevertheless, the accuracy of the measured performance depends on the size of the validation set, which is relatively small (27 lesions), and thus restricts the generalisability of the method. To increase the confidence of the results, a larger ground truth database should be tested (i.e. at least 40 lesions).

Another limitation of the method is the small variety of cancers on which the classifier was trained and tested (mass enhancement and DCIS). A larger variety of cancers (e.g. necrotic core, LCIS) should be tested to check the robustness of the proposed classification system.

# Opportunities and future directions

### Efficient fitting of the proposed empiric model of enhancement

Fitting a nonlinear model of contrast enhancement to every voxel in a DCE-MRI study is time-consuming and requires extensive computing power. The simple form of the proposed model may suggest that using a dedicated non-linear fitting (e.g. Jacobian matrix, adaptive selection of starting values, and so on), will reduce the fitting time dramatically, because it will reduce the number of iterations that the optimisation process requires to converge.

### Visualising empiric model parameters

The proposed empiric model of enhancement has three free parameters. It is possible that a proper conversion of the parameters' values from the fitted empiric model into visual representation may reveal properties about the investigated breast tissue that will be clinically useful (i.e. will help to differentiate malignant from benign more easily) (Vidholm et al., 2007, Mehnert et al., 2005).

### Improve the data reduction rate of the segmentation algorithm

The segmentation algorithm in Chapter 5 reduces the amount of data needing to be interpreted by the radiologist. The segmented amount of tissue can be further reduced by a more careful selection of seed points that can be based on several voxels rather than one. Also, a post-filtering process for the segmented lesions can be applied to eliminate falsely segmented lesions using prior knowledge, such as contextual information (e.g. eliminate the liver from the image based on its location and size or create a mask of the breast volume).

### Voxel-wise classification

The classification system proposed in Chapter 6 is based on two features that can be generated locally (i.e. for a single voxel of a small set of voxel neighbourhood) and do not require prior segmentation of the lesion. This means that the training and the classification can be applied on a voxel-by-voxel basis. The result of such classification can be then visualised, based on the probability for malignancy that the classifier yields for each voxel. This visualisation may provide a good indication for the location of

highly suspicious lesions without the need to segment them in advance. Also, the result of such a classification may be used as the basis for a reliable segmentation of suspicious lesions for further computerised analysis.

# Conclusion

The underlying hypothesis of this research is that **the specificity (and possibly sensitivity) of breast MRI interpretation can be improved** by:

- design features for suspicious lesion classification that will integrate information about tissue kinetic enhancement and morphology
- reducing the subjectivity of breast MRI interpretation by using image analysis and pattern recognition techniques to automatically classify suspicious lesions in the breast.

This thesis has introduced a set of methods for processing and analysing DCE-MR images of the breast. The proposed set of methods can be used as part of a larger CAE system to improve breast cancer diagnosis. The results of this research suggest that by using the methods proposed here, and possibly additional methods (e.g. breast registration, breast volume segmentation), the specificity and sensitivity of DCE-MRI of the breast can indeed be improved (e.g. 93% sensitivity with 77% specificity) by integrating the kinetic and structural information and by using image analysis and pattern recognition techniques.

# Bibliography


ABUTABLEB, A. S. (1989) Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Vision, Graphics, and Image Processing,* 47**,** 22 - 32.

ADAMS, R. & BISCHOF, L. (1994) Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 16**,** 641-647.

AHEARN, T. S., STAFF, R. T., REDPATH, T. W. & SEMPLE, S. I. K. (2005) The use of the Levenberg–Marquardt curve-fitting algorithm in pharmacokinetic modelling of DCE-MRI data. *Physics in Medicine and Biology,* 50**,** N85-N92.

AMERICAN COLLEGE OF RADIOLOGY (2006) Breast imaging reporting and data system (BI-RADS). www.arc.org, American College of Radiology.

ARBACH, L., STOLPEN, A. & REINHARDT, J. M. (2004) Classification of breast MRI lesions using a backpropagation neural network (BNN). *Biomedical Imaging: Macro to Nano, 2004. IEEE International Symposium on,* 1**,** 253-256.

ARIFIN, A. Z. & ASANO, A. (2006) Image segmentation by histogram thresholding using hierarchical cluster analysis. *Pattern Recognition Letters,* 27**,** 1515-1521.

AUSTRALIAN INSTITUTE OF HEALTH AND WELFARE (2007) Cancer in Australia 2006. Australian Institute of Health and Welfare AIHW and Australian Association of Cancer Registries AACR: Canberra.

AUSTRALIAN INSTITUTE OF HEALTH AND WELFARE (2009) Breast cancer in Australia: An Overview, 2009. Canberra, Australian Institute of Health and Welfare and National Breast and Ovarian Cancer Centre, Australian Institute of Health and Welfare.

AWATE, S. & WHITAKER, R. (2005a) Higher-Order Image Statistics for Unsupervised, Information-Theoretic, Adaptive, Image Filtering. *Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society International Conference on.* San Diego, IEEE Computer Society; 1999.

AWATE, S. P. & WHITAKER, R. T. (2005b) Nonparametric Neighborhood Statistics for MRI Denoising. *Lecture Notes in Computer Science,* 3565/2005**,** 12.

AWATE, S. P. & WHITAKER, R. T. (2006) Unsupervised, information-theoretic, adaptive image filtering for image restoration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 28**,** 364-376.

AWATE, S. P. & WHITAKER, R. T. (2007) Feature-Preserving MRI Denoising: A Nonparametric Empirical Bayes Approach. *Medical Imaging, IEEE Transactions on,* 26**,** 1242-1255.

BAO, P. & ZHANG, L. (2003) Noise reduction for magnetic resonance images via adaptive multiscale products thresholding. *IEEE Transactions on Medical Imaging,* 22**,** 1089-1099.

BEHRENS, S., LAUE, H., ALTHAUS, M., BOEHLER, T., KUEMMERLEN, B., HAHN, H. K. & PEITGEN, H. O. (2007) Computer assistance for MR based diagnosis of breast

cancer: Present and future challenges. *Computerized Medical Imaging and Graphics,* 31**,** 236-247.

BLACK, M. J., SAPIRO, G., MARIMONT, D. H. & HEEGER, D. (1998) Robust anisotropic diffusion. *Image Processing, IEEE Transactions on,* 7**,** 421-432.

BRIX, G., SEMMLER, W., PORT, R., SCHAD, L. R., LAYER, G. & LORENZ, W. J. (1991) Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging. *Journal of Computer Assisted Tomography,* 15**,** 621-8.

BUADES, A., COLL, B. & MOREL, J. M. (2004) On image denoising methods. *CMLA Preprint,* 5.

BUADES, A., COLL, B. & MOREL, J. M. (2005a) A non-local algorithm for image denoising. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005.*

BUADES, A., COLL, B. & MOREL, J. M. (2005b) A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation,* 4**,** 490.

BUADES, A., COLL, B. & MOREL, J. M. (2008) Nonlocal Image and Movie Denoising. *International Journal of Computer Vision,* 76**,** 123-139.

BUCKLEY, D. L., KERSLAKE, R. W., BLACKBAND, S. J. & HORSMAN, A. (1994) Quantitative analysis of multi-slice Gd-DTPA enhanced dynamic MR images using an automated simplex minimization procedure. *Magn Reson Med,* 32**,** 646-51.

CHEN, W., GIGER, M. L. & BICK, U. (2006a) A Fuzzy C-Means (FCM)-Based Approach for Computerized Segmentation of Breast Lesions in Dynamic Contrast-Enhanced MR Images. *Academic Radiology,* 13**,** 63-72.

CHEN, W., GIGER, M. L., BICK, U. & NEWSTEAD, G. M. (2006b) Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. *Medical Physics,* 33**,** 2878-2887.

CHEN, W., GIGER, M. L., LAN, L. & BICK, U. (2004) Computerized interpretation of breast MRI: Investigation of enhancement-variance dynamics. *Medical Physics,* 31**,** 1076-1082.

CHEN, X., MICHAEL, B., JONATHAN LOK-CHUEN, L. & NIALL, M. (2005) Simultaneous Segmentation and Registration of Contrast-Enhanced Breast MRI. *Information Processing in Medical Imaging.* Springer Berlin / Heidelberg.

CHENG, H. D., CHEN, Y. H. & JIANG, X. H. (2000) Thresholding using two-dimensional histogram and fuzzy entropy principle. *IEEE Transactions on Image Processing,* 9**,** 732-735.

CLAUS, E. B., RISCH, N., THOMPSON, W. D. & CARTER, D. (1993) Relationship between breast histopathology and family history of breast cancer. *Cancer,* 71**,** 147 - 153.

COIFMAN, R. R. & WICKERHAUSER, M. V. (1994) Adapted waveform analysis as a tool for modeling, feature extraction, and denoising. *Optical Engineering,* 33**,** 2170-2174.

COIFMAN, R. R., WICKERHAUSER, M. V. & WOOG, L. (1997) Adaptive design toolkit software. *New Haven: Fast Mathematical Algorithms and Software Corporation*.

COLLINS, D. J. & PADHANI, A. R. (2004) Dynamic magnetic resonance imaging of tumor perfusion. *Engineering in Medicine and Biology Magazine, IEEE,* 23**,** 65-83.

COMON, P. (1994) Independent component analysis, a new concept. *Signal Processing,* 36**,** 287-314.

CORTES, C. & VAPNIK, V. (1995) Support-vector networks. *Machine learning,* 20**,** 273-297.

COUPE, P., YGER, P., PRIMA, S., HELLIER, P., KERVRANN, C. & BARILLOT, C. (2008) An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. *Medical Imaging, IEEE Transactions on,* 27**,** 425-441.

DANGETI, S. V. (2002) Denoising Techniques-A Comparison. *Electrical and Computer Engineering.* Visakhapatnam, India, Andhra University College of Engineering.

DEGANI, H., GUSIS, V., WEINSTEIN, D., FIELDS, S. & STRANO, S. (1997) Mapping pathophysiological features of breast tumors by MRI at high spatial resolution. *Nat Med,* 3**,** 780-782.

DONOHO, D. L. & JOHNSTONE, I. M. (1994) Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika,* 81**,** 425-455.

FAN, X., MEDVED, M., RIVER, J. N., ZAMORA, M., COROT, C., ROBERT, P., BOURRINET, P., LIPTON, M., CULP, R. M. & KARCZMAR, G. S. (2004) New model for analysis of dynamic contrast-enhanced MRI data distinguishes metastatic from nonmetastatic transplanted rodent prostate tumors. *Magnetic Resonance in Medicine,* 51**,** 487-494.

FISCHER, U. & BRINCK, U. (2004) *Practical MR Mammography,* Stuttgart, Germany, Thieme Publishing Group.

FURMAN-HARAN, E. & DEGANI, H. (2002) Parametric analysis of breast MRI. *Journal of Computer Assisted Tomography,* 26**,** 376-86.

GAL, Y., MEHNERT, A., BRADLEY, A., KENNEDY, D. & CROZIER, S. (2009a) Feature and Classifier Selection for Automatic Classification of Lesions in Dynamic Contrast-Enhanced MRI of the breast. *Proceedings Digital Image Computing: Techniques and Applications (DICTA).* Melbourne, Australia.

GAL, Y., MEHNERT, A., BRADLEY, A., MCMAHON, K. & CROZIER, S. (2007a) Automatic Segmentation of Enhancing Breast Tissue in Dynamic Contrast-Enhanced MR Images. IN J. BOTTEMA, M., MAEDER, A., REDDING, N. & VAN DEN HENGEL, A. (Eds.) *Proceedings Digital Image Computing: Techniques and Applications (DICTA).* Adelaide, Australia.

GAL, Y., MEHNERT, A., BRADLEY, A., MCMAHON, K. & CROZIER, S. (2007b) An evaluation of four parametric models of contrast enhancement for dynamic magnetic resonance imaging of the breast. *IEEE Engineering in Medicine and Biology Society (EMBC).* Lyon, France.

GAL, Y., MEHNERT, A., BRADLEY, A., MCMAHON, K. & CROZIER, S. (2009b) A Variation on Non-Local Means for the Denoising of Dynamic Contrast-Enhanced MR Images. *Symposium on GPGPU Techniques for Medical Image Processing and Simulation.* Brisbane, Australia, EMB.

GAL, Y., MEHNERT, A., BRADLEY, A., MCMAHON, K., KENNEDY, D. & CROZIER, S. (2008) A new denoising method for dynamic contrast-enhanced MRI. *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE.* Vancouver, British Columbia, Canada.

GAL, Y., MEHNERT, A., BRADLEY, A., MCMAHON, K., KENNEDY, D. & CROZIER, S. (2010) Denoising of Dynamic Contrast-Enhanced MR Images Using Dynamic Non-Local Means. *IEEE Transactions on Medical Imaging,* 29**,** 302-310.

GEMAN, S., BIENENSTOCK, E. & DOURSAT, R. (1992) Neural networks and the bias/variance dilemma. *Neural computation,* 4**,** 1-58.

GERIG, G., KUBLER, O., KIKINIS, R. & JOLESZ, F. A. (1992) Nonlinear anisotropic filtering of MRI data. *Medical Imaging, IEEE Transactions on,* 11**,** 221-232.

GILHUIJS, K. G., GIGER, M. L. & BICK, U. (1999) A method for computerized assessment of tumor extent in contrast-enhanced MR images of the breast. *Computer-aided diagnosis in medical imaging.* Philadelphia, PA, Elsevier Science B.V.

GONZALEZ & WOODS, R. E. (2002) *Digital Image Processing*, Prentice Hall.

GUDBJARTSSON, H. & PATZ, S. (1996) The Rician distribution of noisy MRI data (vol 34, pg 910, 1995). *Magnetic Resonance in Medicine,* 36**,** 332-332.

HAACKE, E., THOMPSON, M., BROWN, R. & VENKATESAN, R. (1999) *Magnetic resonance imaging: physical principles and sequence design*, Wiley New York.

HANLEY, J. A. & MCNEIL, B. (1982) The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology,* 143**,** 29-36.

HARALICK, R., DINSTEIN, I. & SHANMUGAM, K. (1973) Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics,* 3**,** 610-621.

HAYTON, P., BRADY, M., TARASSENKO, L. & MOORE, N. (1997) Analysis of dynamic MR breast images using a model of contrast enhancement. *Medical Image Analysis,* 1**,** 207-24.

HAYTON, P. M. (1998) Analysis of contrast-enhanced breast MRI. *Department of Engineering Science.* Oxford, The University of Oxford, Oxford, UK.

HEIBERG, E. V., PERMAN, W. H., HERRMANN, V. M. & JANNEY, C. G. (1996) Dynamic sequential 3D gadolinium-enhanced MRI of the whole breast. *Magnetic Resonance Imaging,* 14**,** 337-348.

HITTMAIR, K., GOMISCEK, G., LANGENBERGER, K., RECHT, M., IMHOF, H. & KRAMER, J. (1994) Method For the Quantitative Assessment of Contrast Agent Uptake in Dynamic Contrast-Enhanced Mrt. *Magnetic Resonance in Medicine,* 31**,** 567-571.

JACKSON, A., BUCKLEY, D. L. & PARKER, G. J. M. (2005) *Dynamic contrast-enhanced magnetic resonance imaging in oncology* New York, Springer.

JACOBS, M. A., BARKER, P. B., BLUEMKE, D. A., MARANTO, C., ARNOLD, C., HERSKOVITS, E. H. & BHUJWALLA, Z. (2003) Benign and Malignant Breast Lesions: Diagnosis with Multiparametric MR Imaging. *Radiology,* 229**,** 225-232.

JAIN, A. K., DUIN, R. P. W. & MAO, J. (2000) Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI),* 22**,** 4–37.

JANSEN, S. A., FAN, X., KARCZMAR, G. S., ABE, H., SCHMIDT, R. A. & NEWSTEAD, G. M. (2008) Differentiation between benign and malignant breast lesions detected by bilateral dynamic contrast-enhanced MRI: A sensitivity and specificity study. *Magnetic Resonance in Medicine,* 59**,** 747-754.

JIANG, W., BAKER, M. L., WU, Q., BAJAJ, C. & CHIU, W. (2003) Applications of a bilateral denoising filter in biological electron microscopy. *Journal of Structural Biology,* 144**,** 114-122.

KASS, M., WITKIN, A. & TERZOPOULOS, D. (1988) Snakes: Active contour models. *International Journal of Computer Vision,* 1**,** 321-331.

KEREN, D., OSADCHY, M. & GOTSMAN, C. (2000) Anti-Faces for Detection. *Proceedings of the 6th European Conference on Computer Vision-Part I***,** 134-148.

KEREN, D., OSADCHY, M. & GOTSMAN, C. (2001) Antifaces: a novel, fast method for image detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 23**,** 747-761.

KERVRANN, C., BOULANGER, J. & COUPE, P. (2007) Bayesian Non-local Means Filter, Image Redundancy and Adaptive Dictionaries for Noise Removal. *LECTURE NOTES IN COMPUTER SCIENCE,* 4485**,** 520-532.

KOHLER, T. & LORENZ, D. (2005) A comparison of denoising methods for one dimensional time series. Bremen, Germany, University of Bremen.

KUHL, C. K., MIELCARECK, P., KLASCHIK, S., LEUTNER, C., WARDELMANN, E., GIESEKE, J. & SCHILD, H. H. (1999) Dynamic Breast MR Imaging: Are Signal Intensity Time Course Data Useful for Differential Diagnosis of Enhancing Lesions? *Radiology,* 211**,** 101-110.

KUPINSKI, M. A. & GIGER, M. L. (1998) Automated Seeded Lesion Segmentation on Digital Mammograms. *IEEE TRANSACTIONS ON MEDICAL IMAGING,* 17**,** 510-517.

LANDINI, G. & RIPPIN, J. W. (1996) How Important is Tumor Shape? Quantification of the Epithelial–Connective Tissue Interface in Oral Lesions Using Local COonnected Fractal Dimension Analysis. *The Journal of Pathology,* 179**,** 210-217.

LARSSON, H. B., STUBGAARD, M., FREDERIKSEN, J. L., JENSEN, M., HENRIKSEN, O. & PAULSON, O. B. (1990) Quantitation of blood-brain barrier defect by magnetic resonance imaging and gadolinium-DTPA in patients with multiple sclerosis and brain tumors. *Magnetic Resonance Medicine,* 16**,** 117-31.

LEE, S. H., KIM, J. H., KIM, K. G., PARK, J. S., PARK, S. J. & MOON, W. K. (2007) Optimal Clustering of Kinetic Patterns on Malignant Breast Lesions: Comparison between K-means Clustering and Three-time-points Method in Dynamic Contrast-enhanced MRI. *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE.*

LEGGETT, J. (2004) Multi-layer Gradient Coil Design. *School of Physics & Astronomy, Faculty of Science.* Nottingham, UK, University of Nottingham.

LEVIN, A., ZOMET, A. & WEISS, Y. (2003) Learning how to inpaint from global image statistics. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on.*

LI, H. D., KALLERGI, M., CLARKE, L. P., JAIN, V. K. & CLARK, R. A. (1995) Markov random field for tumor detection in digital mammography. *Medical Imaging, IEEE Transactions on,* 14**,** 565-576.

LIU, P. F., DEBATIN, J. F., CADUFF, R. F., KACL, G., GARZOLI, E. & KRESTIN, G. P. (1998) Improved diagnostic accuracy in dynamic contrast enhanced MRI of the breast by combined quantitative and qualitative analysis. *British Journal of Radiology,* 71**,** 501-509.

LIU, Y. L., WANG, J., CHEN, X., GUO, Y. W. & PENG, Q. S. (2008) A Robust and Fast Non-Local Means Algorithm for Image Denoising. *Journal of Computer Science and Technology,* 23**,** 270-279.

LUCAS-QUESADA, F. A., SINHA, U. & SINHA, S. (2005) Segmentation strategies for breast tumors from dynamic MR images. *Journal of Magnetic Resonance Imaging,* 6**,** 753 - 763.

LYSAKER, M., LUNDERVOLD, A. & XUE-CHENG, T. (2003) Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *Image Processing, IEEE Transactions on,* 12**,** 1579-1590.

MACOVSKI, A. (1996) Noise in MRI. *Magn Reson Med,* 36**,** 494-7.

MAHMOUDI, M. & SAPIRO, G. (2005) Fast image and video denoising via nonlocal means of similar neighborhoods. *Signal Processing Letters, IEEE,* 12**,** 839-842.

MALLADI, R. & SETHIAN, J. A. (1995) Level set methods for curvature flow, image enhancement, and shape recovery in medical images. *Proc. of Conf. on Visualization and Mathematics, June.*

MALLADI, R. & SETHIAN, J. A. (1996) Level set and fast marching methods in image processing and computer vision. *Image Processing, 1996. Proceedings.*

MALLAT, S. & ZHONG, S. (1992) Characterization of signals from multiscale edges. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 14**,** 710-732.

MANJÓN, J. V., CARBONELL-CABALLERO, J., LULL, J. J., GARCÍA-MARTÍ, G., MARTÍ-BONMATÍ, L. & ROBLES, M. (2008) MRI denoising using Non-Local Means. *Medical Image Analysis,* 12**,** 514-523.

MANJON, J. V., ROBLES, M. & THACKER, N. A. (2007) Multispectral MRI De-noising Using Non-Local Means. *Medical Image Understanding and Analysis (MIUA).*

MARQUARDT, D. W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics,* 11**,** 431-441.

MARRON, J. S. & TSYBAKOV, A. B. (1995) Visual Error Criteria for Qualitative Smoothing. *Journal of the American Statistical Association,* 90**,** 499-507.

MARTEL, A. L. (2006) A Fast Method of Generating Pharmacokinetic Maps from Dynamic Contrast-Enhanced Images of the Breast. *MICCAI.*

MATHWORKS (2007) MATLAB. 7.4.0.287 (R2007a) ed., The MathWorks inc.

MEHNERT, A., BENGTSSON, E., MCMAHON, K., KENNEDY, D., WILSON, S. & CROZIER, S. (2005) Visualisation of the pattern of contrast enhancement in dynamic breast MRI. *Proceedings WDIC2005, APRS Workshop on Digital Image Computing* Griffith University, Southbank, Brisbane, Australia, The University of Queensland.

MEHNERT, A. & JACKWAY, P. (1997) An improved seeded region growing algorithm. *Pattern Recognition Letters,* 18**,** 1065-1071.

MOORE, B. (1981) Principal component analysis in linear systems: Controllability, observability, and model reduction. *Automatic Control, IEEE Transactions on,* 26**,** 17-32.

MORRIS, E. A. & LIBERMAN, L. (2005) *Breast MRI - diagnosis and intervention,* New York, Springer.

MURASE, K., YAMAZAKI, Y., SHINOHARA, M., KAWAKAMI, K., KIKUCHI, K., MIKI, H., MOCHIZUKI, T. & IKEZOE, J. (2001) An anisotropic diffusion method for denoising dynamic susceptibility contrast-enhanced magnetic resonance images. *Physics in Medicine and Biology,* 46**,** 2713-2723.

NATTKEMPER, T. W., ARNRICH, B., LICHTE, O., TIMM, W., DEGENHARD, A., POINTON, L., HAYES, C. & LEACH, M. O. (2005) Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods. *Artificial Intelligence in Medicine,* 34**,** 129-139.

NELDER, J. A. & MEAD, R. (1965) A simplex method for function minimization. *Computer Journal,* 7**,** 308-313.

NOWAK, R. D. (1999) Wavelet-based Rician noise removal for magnetic resonance imaging. *Image Processing, IEEE Transactions on,* 8**,** 1408-1419.

OREL, S. & SCHNALL, M. (2001) MR Imaging of the Breast for the Detection, Diagnosis, and Staging of Breast Cancer1. *Radiology,* 220**,** 13.

OTSU, N. (1979) A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on System, Man and Cybernetics,* SMC-9**,** 62-66.

PAL, N. R. & PAL, S. K. (1993) A review on image segmentation techniques. *Pattern Recognition,* 26**,** 1277-1294.

PAN, Q., ZHANG, L., DAI, G. & ZHANG, H. (1999) Two denoising methods by wavelet transform. *Signal Processing, IEEE Transactions on,* 47**,** 3401-3406.

PERONA, P. & MALIK, J. (1990) Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 12**,** 629-639.

PETRICK, N., CHAN, H. P., SAHINER, B. & HELVIE, M. A. (1999) Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. *Medical Physics,* 26**,** 1642.

PIZURICA, A., PHILIPS, W., LEMAHIEU, I. & ACHEROY, M. (2003) A versatile wavelet domain noise filtration technique for medical imaging. *Medical Imaging, IEEE Transactions on,* 22**,** 323-331.

POHLE, R. & TOENNIES, K. D. (2001) Segmentation of medical images using adaptive region growing. *Proc. SPIE Medical Imaging,* 4322**,** 1337–1346.

POOLE, M. (2007) Improved Equipment and Techniques for Dynamic Shimming in High Field MRI. Nottingham, University of Nottingham.

REVOL, C. & JOURLIN, M. (1997) A new minimum variance region growing algorithm for image segmentation. *Pattern Recognition Letters,* 18**,** 249-258.

RICE, S. (1944) Mathematical analysis of random noise. *Bell Systems Technology Journal,* 23**,** 282-332.

RIMEY, R. D. & COHEN, F. S. (1988) A maximum-likelihood approach to segmenting range data. *Robotics and Automation, IEEE Journal of [see also IEEE Transactions on Robotics and Automation],* 4**,** 277-286.

RODENACKER, K. & BENGTSSON, E. (2003) A feature set for cytometry on digitized microscopic images. *Analytical Cellular Pathology,* 25**,** 1-36.

ROSSET, A., SPADOLA, L. & RATIB, O. (2004) OsiriX: an open-source software for navigating in multidimensional DICOM images. *Journal of Digital Imaging,* 17**,** 205-216.

SETHIAN, J. A. (1996) A fast marching level set method for monotonically advancing fronts.

SETHIAN, J. A. (1999) Level Set Methods and Fast Marching Methods. *Evolving interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science. Univ. Press, Cambridge*.

SEZGIN, M. & SANKUR, B. (2004) Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging,* 13**,** 146-168.

SINHA, S., LUCAS-QUESADA, F. A., DEBRUHL, N. D., SAYRE, J., FARRIA, D., GORCZYCA, D. P. & BASSETT, L. W. (1997) Multifeature analysis of Gd-enhanced MR images of breast lesions. *Journal of Magnetic Resonance Imaging,* 7**,** 1016-26.

SNYDER, W. E. & QI, H. (2004) *Machine Vision,* Cambridge, UK, Cambridge University Press.

SYKULSKI, J. K., ROTARU, M., SABENE, M. & SANTILLI, M. (1998) Comparison of optimization techniques for electromagnetic applications. *COMPEL – The International Journal for Computation and Mathematics in Electrical and Electronic Engineering,* 17**,** 171-176.

SZABO, B. K., ASPELIN, P., KRISTOFFERSEN WIBERG, M. & BONÉ, B. (2003) Dynamic MR Imaging of the breast: Analysis of kinetic and morphologic diagnostic criteria. *Acta Radiologica,* 44**,** 379-386.

TISDALL, D. & ATKINS, M. S. (2005) MRI denoising via phase error estimation. IN FITZPATRICK, M. J. (Ed.) *Procedings of SPIE.* San Diego, CA, SPIE, Bellibham, WA.

TOFTS, P. S., BERKOWITZ, B. & SCHNALL, M. D. (1995) Quantitative-analysis of dynamic Gd-DTPA enhancement in breast-tumors using a permeability model. *Magnetic Resonance in Medicine,* 33**,** 564-568.

TOFTS, P. S. & BERKOWITZ, B. A. (1994) Measurement of capillary permeability from the Gd enhancement curve: A comparison of bolus and constant infusion injection methods. *Magnetic Resonance Imaging,* 12**,** 81-91.

TOFTS, P. S. & KERMODE, A. G. (1991) Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magnetic Resonance Medicine,* 17**,** 357-67.

TOMASI, C. & MANDUCHI, R. (1998) Bilateral filtering for gray and color images. *Proceedings of the Sixth International Conference on Computer Vision.* Bombay, India.

TORHEIM, G., GODTLIEBSEN, F., AXELSON, D., KVISTAD, K. A., HARALDSETH, A. & RINCK, P. A. (2001) Feature extraction and classification of dynamic contrast-enhancedT2*-weighted breast image data. *IEEE Transactions on Medical Imaging,* 20**,** 1293-1301.

TURK, M. A. & PENTLAND, A. P. (1991) Face recognition using eigenfaces. *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on.*

VIDHOLM, E., MEHNERT, A., BENGTSSON, E., WILDERMOTH, M., MCMAHON, K., WILSON, S. & CROZIER, S. (2007) Hardware-accelerated volume visualisation of parametrically mapped dynamic breast MRI data. *The 10th International Conference on Medical Image Computing and Computer Assisted Intervention* Brisbane, Australia MICCAI.

WALKER, S. A., MILLER, D. & TANABE, J. (2006) Bilateral spatial filtering: Refining methods for localizing brain activation in the presence of parenchymal abnormalities. *NeuroImage,* 33**,** 564-569.

WARREN, R. & COULTHARD, A. (2002) *Breast MRI in practice,* London, UK, Martin Dunitz.

WARREN, R., POINTON, L., THOMPSON, D., HOFF, R., GILBERT, F., PADHANI, A., EASTON, D., LAKHANI, S. & LEACH, M. (2005) Reading Protocol for Dynamic Contrast-enhanced MR Images of the Breast: Sensitivity and Specificity Analysis1. *Radiology,* 236**,** 779.

WEBB, A. (2003) *Statistical Pattern Recognition,* Chichester, England, John Wiley & Sons Ltd.

WIEST-DAESSLE, N., PRIMA, S., COUPE, P., MORRISSEY, S. P. & BARILLOT, C. (2008) Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Application to DT-MRI. *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI).* New York.

WIEST-DAESSLE, N., PRIMA, S., COUPE, P., PATRICK MORRISSEY, S. & BARILLOT, C. (2007) Non-Local Means Variants for Denoising of Diffusion-Weighted and Diffusion Tensor MRI. IN AYACHE, N., OURSELIN, S. & MAEDER, A. (Eds.) *Medical Image Computing and Computer-Assisted Intervention -- MICCAI.* Brisbane, Australia, Springer.

WONG, A. K. C. & SAHOO, P. K. (1989) A gray-level threshold selection method based on maximum entropyprinciple. *Systems, Man and Cybernetics, IEEE Transactions on,* 19**,** 866-871.

WOOD, J. C. & JOHNSON, K. M. (1999) Wavelet packet denoising of magnetic resonance images: Importance of Rician noise at low SNR. *Magnetic Resonance in Medicine,* 41**,** 631-635.

XIE, J., HENG, P.-A., HO, S. S. M. & SHAH, M. (2006) Image Diffusion Using Saliency Bilateral Filter. *Medical Image Computing and Computer-Assisted Intervention (MICCAI).* Copenhagen, Denmark, Springer Berlin / Heidelberg.

XUAN, J., TULAY, A. & WANG, Y. (1995) Segmentation of Magnetic Resonance Brain Image: Integrating Region Growing and Edge Detection. *Image Processing, 1995. Proceedings., International Conference on,* 3**,** 544-547.

# Appendix A:
# Entropy based thresholding

Entropy based thresholding methods determine a threshold that maximizes the sum of the information below and above the threshold (Abutableb, 1989, Wong and Sahoo, 1989, Sezgin and Sankur, 2004). They are attractive because they are fully automatic and do not assume any particular distributional model of the intensity levels in the image. Also, they can incorporate one or more features (e.g. similarity to neighbouring pixels) into the threshold decision process and thus allow better control over specificity and sensitivity. Breast lesions in MRI are complicated in nature and have a high variety in appearance (Morris and Liberman, 2005). The entropy thresholding was thus chosen to be used as a first stage segmentation of suspicious lesions in DCE-MRI of the breast, because it is robust to the distribution of intensities in the segmented object in addition to allowing it to create a synergy between two (or more) features in the thresholding process.

In its simplest form, the binarization threshold minimizes the sum of Shannon's entropy for the foreground and background intensity histograms. The method is based on principals from information theory; a description of the method follows (Abutableb, 1989, Sezgin and Sankur, 2004).

Given a set (alphabet) with N letters, $A = \{a_i\}_{i=1}^{N}$, represents all the possible letters (i.e. the intensity levels in an image) that are used in a given information channel (i.e. an image), the probability of appearance for each letter is defined as follows:

$$p_i = P(a_i|A),$$

where $\sum_{i=1}^{N} P_i = 1$

The amount of information that a letter, $a_i$, introduces in the system is defined as

$$I(a_i) \coloneqq \log_N \frac{1}{p_i} = -\log_N(p_i)$$

The quantity $I(a_i)$ is often called the self-information of $a_i$.

Thus, the average amount of information in the channel (image) is:

$$H_N = -\sum_{i=1}^{N} p_i \log(p_i)$$

Now, in the case of an image with histogram, $\overline{H} = \{h_i\}_{i=1}^{M}$, we normalise the histogram to become a probability density function (PDF), where each intensity bin in the histogram represents the probability of a randomly-selected pixel to have the corresponding intensity, *i*. In this case the corresponding probability $p_i$ will be:

$$p_i = \frac{h_i}{\sum_{m=1}^{M} h_m}$$

where *M* is the number of intensity levels in the histogram.

We will now define the amount of information below the threshold as:

$$H(A) = -\sum_{i=1}^{s} \left(\frac{p_i}{P_s}\right) \log\left(\frac{p_i}{P_s}\right)$$

where

$$P_s = \sum_{i=1}^{s} p$$

We now define:

$$H_s = -\sum_{i=1}^{s} p_i \log(p_i)$$

Using (5.7), we can now write equation (5.5) as follows:

$$H(A) = -(\frac{1}{P_s}) \sum_{i=1}^{s} [p_i \log(p_i) - p_i \log(P_s)] = \log(P_s) + \frac{H_s}{P_s}$$

The information above the threshold will be defined as:

$$H(B) = \sum_{i=s+1}^{M} \left(\frac{p_i}{1-P_s}\right) \log\left(\frac{p_i}{1-P_s}\right)$$

Equation (5.9) can be written as:

$$H(B) = -\left(\frac{1}{1-P_s}\right) \sum_{i=s+1}^{M} p_i [\log(p_i) - \log(1-P_s)] = \log(1-P_s) + \frac{H_m - H_s}{1-P_s}$$

where $H_m = -\sum_{i=1}^{M} p_i \log(p_i)$

The aim is to find a threshold *s* that maximizes the entropy criterion:

$$\psi = H(A) + H(B) = \log[P_s(1-P_s)] + \frac{H_s}{P_s} + \frac{H_m - H_s}{1-P_s}$$

The threshold, *s*, which maximizes $\psi$ is used to differentiate between the objects of interest and the background of the image.

In the two-dimensional entropy method, the aim is to find two thresholds *s, t*, that maximize the same basic criterion. The histogram in this case is two-dimensional and one of its axes is usually the intensity value of the pixel, while the other axis represents a feature of the object that may introduce additional relevant information, such as spatial information (e.g. the histogram of a filtered version of the image) (Cheng et al., 2000, Abutableb, 1989).

In the two dimensional case we define:

$$P_{st} = \sum_{i=1}^{s} \sum_{j=1}^{t} p_{ij}$$

and

$$H_{st} = \sum_{i=1}^{s} \sum_{j=1}^{t} p_{ij} \log(p_{ij})$$

The two-dimensional entropy criterion will then be defined as:

$$\psi(s,t) = H(A) + H(B) = \log[P_{st}(1 - P_{st})] + \frac{H_{st}}{P_{st}} + \frac{H_{mm} - H_{st}}{1 - P_{st}}$$

The foreground pixels will be those that satisfy both thresholds and the rest will be considered as background.

Other entropy-based methods consider different parameterizations of entropies over the image histogram, such as the Kullback-Leibler distance and cross-entropy between the original and binarized images (Sezgin and Sankur, 2004).

# Appendix B: Seeded Region Growing Segmentation

Of the region-based segmentation techniques, one family is region growing segmentation. This family of methods usually requires starting points or "seeds" from which the growing process will start (Adams and Bischof, 1994, Petrick et al., 1999, Mehnert and Jackway, 1997, Revol and Jourlin, 1997, Xuan et al., 1995, Pohle and Toennies, 2001, Kupinski and Giger, 1998).

The process of seeded region growing (SRG) starts with a number of seeds that have been grouped into $n$ sets of connected components, say, $A_1, A_2, \ldots, A_n$. Sometimes, individual sets will contain a single point. The choice of seeds depends on the features of interest and what is relevant or noise. The process starts from the seeds and evolves iteratively. Each step of the algorithm adds a new pixel to one of the initial sets, according to a predefined similarity criterion.

Let us define $S$ as the set of all the unvisited pixels that border at least one of the growing regions:

$$S = \left\{ x \notin \bigcup_{i=1}^{n} A_i \,\middle|\, N(x) \cap \bigcup_{i=1}^{n} A_i = \phi \right\}$$

where *N(x)* is the set of immediate neighbours of the pixel *x*. If for $x \in S$, *N(x)* meets just one of the $A_i$, then the predefined metric, $\delta(x, A_i)$, which is computed to measure how different *x* is from the region $A_i$ it adjoins, is used to determine if *x* should be joined to $A_i$. The simplest definition for $\delta(x, A_i)$ is:

$$\delta(x, A_i) = |g(x) - \langle A_i \rangle|$$

where *g(x)* is the intensity of the pixel *x* and $\langle A_i \rangle$ is the mean intensity of the pixels of the region $A_i$. If *N(x)* meets two or more of the $A_i$, then *x* is assigned to the $A_i$ that minimizes $\delta(x, A_i)$ (Adams and Bischof, 1994). The special case of a tie between two $A_i$ should be considered and may be handled, if needed, by introducing additional assignment criteria or by a random selection.

The growing criterion can be based on more than one measurement. In fact, the simplest measurement will most likely cause the region to grow beyond the desired boundaries. A stopping criterion, based on additional information, such as the order of magnitude of the gradient, can be incorporated into the growing criterion (Xuan et al., 1995). After the growing process is complete, a merging stage may be considered. In this stage, adjacent regions are merged according to a similarity criterion. Jianhua *et al.* (Xuan et al., 1995) propose a similarity criterion that is based on the intensity statistics:

$$S_{sim}(A, B) = \frac{|\mu_A - \mu_B|}{\max(1, \sigma_A + \sigma_B)}$$

where $\mu$ is the mean intensity and $\sigma$ is the standard deviation. The maximum in the denominator ensures that the result is stable in cases where the standard deviation is very small.

Revol and Jourlin (Revol and Jourlin, 1997) propose a minimum variance criterion on the growing process, yielding a homogenous region that does not necessarily contain the original seeds. Their algorithm may add new pixels to a region in one step and eliminate certain pixels from a region in a following sub-step if they decrease the homogeneity of the region.

# Appendix C:

# MATLAB code for NL-Means and DNLM algorithms

```
function fimg = nlmeans4d(img, options)

% fimg = nlmeans4d(img, options)
%
% img is a 4D image
% options is a structure containing:
%  options.k is a vector, holding the radius of the comparison window for each dimension
%              if a single value is used an isotropic window will be used
%  options.sig is the estimated noise standard deviation. Default: automatic estimation.
%  options.beta is the denoising factor. If not given, it will be set to be 1
%              note that h^2 = 2 * beta * sig^2 * n_of_neighbours (according to the similarity window size)
%  options.dstsig is the std, in pixels, to use for distance weighting (default is max(k)/2)
%  options.win is a vector, holding the size of search radius in each direction if a single value is used an isotropic search
%              area will be used
%  options.ancor is a scalar determining the method to be used:
%              0 (default) for old DNLM (EMBC 2008)
%              1 for classic NL-Means,
%              2 New DNLM (IEEE TMI 2009)

size_x = size(img,1);
size_y = size(img,2);
size_z = size(img,3);
size_w = size(img,4);

fimg = zeros(size_x, size_y, size_z, size_w);

if length(options.k) == 1
   k = [options.k options.k options.k options.k];
else
   k = options.k;
end

if isfield(options, 'dstsig')
   dst_sig2 = options.dstsig * options.dstsig;
else
   dst_sig2 = max(k)*max(k)/4;
end

if isfield(options, 'ancor')
   ancor = options.ancor; % 0=Old DNLM (EMBC 2008), 1=ENLM (3D NL-Means), 2= DNLM (IEEE TMI 2009)
else
   ancor = 0; % old DNLM
end

if isfield(options, 'win')
   if length(options.win) == 1
      win = [options.win options.win options.win options.win];
   else
      win = options.win;
   end
else
   win = [0 0 0 size(img,4)];
```

```
end

if isfield(options, 'sig')
   sig = options.sig;
else
   est_noise = 0;
   est_count = 0;
   for w=1:size_w
      for z=1:size_z
         [est_sig est_SNR est_MAV] = estimate_noise_dwt(img(:,:,z,w));
         est_noise = est_noise + est_MAV;
         est_count = est_count + 1;
      end
   end
   sig = est_noise / est_count;
end

if isfield(options, 'beta')
   sig2 = 2 * options.beta * sig * sig * (prod(2*k+1)-1); % = 2 * beta * noise^2 * n_of_neighbours
else
   sig2 = 2 * sig * sig * (prod(2*k+1)-1); % = 2 * noise^2 * n_of_neighbours
end

% Zero padding
size_px = size_x + 2 * k(1);
size_py = size_y + 2 * k(2);
size_pz = size_z + 2 * k(3);
size_pw = size_w + 2 * k(4);
pimg = zeros(size_px, size_py, size_pz, size_pw);
pimg((k(1)+1):(k(1)+size_x),(k(2)+1):(k(2)+size_y),(k(3)+1):(k(3)+size_z),(k(4)+1):(k(4)+size_w))=img;

% Building a distance Gaussian filter
dist_flt = zeros(k(1)*2+1, k(2)*2+1, k(3)*2+1, k(4)*2+1);
for wx=1:(k(1)*2+1)
   for wy=1:(k(2)*2+1)
      for wz=1:(k(3)*2+1)
         for ww=1:(k(4)*2+1)
            dst = norm([wx-k(1)-1 wy-k(2)-1 wz-k(3)-1 ww-k(4)-1]);
            dist_flt(wx,wy,wz,ww)=exp(-dst*dst/dst_sig2);
         end
      end
   end
end
dist_flt(k(1)+1,k(2)+1,k(3)+1,k(4)+1)=dist_flt(k(1),k(2)+1,k(3)+1,k(4)+1); % Reduce central weight to avoid over-weighting
dist_flt = dist_flt / sum(dist_flt(:));

% pixel external loops
for w=(k(4)+1):(k(4)+size_w)
   for z=(k(3)+1):(k(3)+size_z)
      for y=(k(2)+1):(k(2)+size_y)
         for x=(k(1)+1):(k(1)+size_x)
            pix_win = pimg((x-k(1)):(x+k(1)), (y-k(2)):(y+k(2)), (z-k(3)):(z+k(3)), (w-k(4)):(w+k(4)));
            mean_pix_win = mean(pix_win(:));
            win_len = length(pix_win(:));
            Zi = 0;
            Wval_sum = 0;

            % internal search loops
            for ww=max(w-win(4),k(4)+1):min(w+win(4),k(4)+size_w)
               for zz=max(z-win(3),k(3)+1):min(z+win(3),k(3)+size_z)
                  for yy=max(y-win(2),k(2)+1):min(y+win(2),k(2)+size_y)
                     for xx=max(x-win(1),k(1)+1):min(x+win(1),k(1)+size_x)
                        comp_win = pimg((xx-k(1)):(xx+k(1)), (yy-k(2)):(yy+k(2)), (zz-k(3)):(zz+k(3)), (ww-k(4)):(ww+k(4)));
                        mean_comp_win = mean(comp_win(:));
                        if (mean_comp_win > 0)
                           if ww ~= w && ancor ~= 1
                              diff_win = comp_win-pix_win;
                              diff_mean = sum(abs(diff_win(:)))/win_len;
                              if diff_mean > sig || ancor == 0 % Old DNLM
                                 comp_win_factor = mean_pix_win / mean_comp_win; % original factor
                                 comp_win = comp_win * comp_win_factor;
                              else
                                 comp_win_factor = 1;
                              end
                           else
                              comp_win_factor = 1;
```

```
                    end

                    diff_win = comp_win-pix_win;
                    diff_nrm2 = sum(dist_flt(:).*diff_win(:).*diff_win(:)); % norm2 convolved with a Gaussian
                    nb_Wij = exp(-diff_nrm2/sig2);

                    Zi = Zi + nb_Wij;
                    Wval_sum = Wval_sum + (nb_Wij*pimg(xx,yy,zz,ww)*comp_win_factor);
                  end
                end % closing internal search loops
              end % closing internal search loops
            end % closing internal search loops
          end % closing internal search loops

          if (Zi > 0)
            fimg(x-k(1),y-k(2),z-k(3),w-k(4)) = Wval_sum / Zi; % normalising
          else
            fimg(x-k(1),y-k(2),z-k(3),w-k(4)) = 0; % normalising
          end

        end % closing pixel external loops
      end % closing pixel external loops
    end % closing pixel external loops
end
```

# Appendix D:

# The ACR-BIRADS Lexicon

**ACR BI-RADS®–MRI Lexicon Classification Form**

For each of the following categories, select the term that best describes the dominant lesion feature. Wherever possible, definitions and descriptions used in BI-RADS® for mammography will be applied to MRI of the breast. This form is for data collection and does not constitute a written MRI report.

**LESION TYPE (select one)**

**A. Focus/Foci (tiny spot of enhancement, < 5 mm) if only finding, *GO TO SECTION E***

**B. Mass (three-dimensional space-occupying lesion that is one process, usually round, oval, or irregular in shape).**

| **Shape** *(select one)* | **Description** |
|---|---|
| ❑ Round | Spherical or ball-shaped |
| ❑ Oval | Elliptical or egg-shaped |
| ❑ Lobular | Undulating contour |
| ❑ Irregular | Uneven shape (not round, oval, or lobulated) |

| **Margin** *(select one)* | **Description** |
|---|---|
| ❑ Smooth | Well-circumscribed and well-defined margin |
| ❑ Irregular | Uneven margin can be round or jagged (not smooth or spiculated) |
| ❑ Spiculated | Characterized by radiating lines |

| **Mass Enhancement** *(select one)* | **Description** |
|---|---|
| ❑ Homogeneous | Confluent uniform enhancement |
| ❑ Heterogeneous | Nonspecific mixed enhancement |
| ❑ Rim enhancement | Enhancement more pronounced at the periphery of mass |
| ❑ Dark internal septation | Dark nonenhancing lines within a mass |
| ❑ Enhancing internal septation | Enhancing lines within a mass |
| ❑ Central enhancement | Enhancement more pronounced at center of mass |

**C. Non-Mass-Like Enhancement (in an area that is not a mass)**

| **Distribution Modifiers** *(select one)* | **Description** |
|---|---|
| ❑ Focal area | Enhancement in a confined area, less than 25% of quadrant |
| ❑ Linear | Enhancement in a line that may not conform to a duct |
| ❑ Ductal | Enhancement in a line that may have branching, conforming to a duct |

**C. Non-Mass-Like Enhancement (in an area that is not a mass) (continued)**

| **Distribution Modifiers** *(select one)* | **Description** |
|---|---|
| ❑ Segmental | Triangular region of enhancement, apex pointing to nipple, suggesting a duct or its branches |
| ❑ Regional | Enhancement in a large volume of tissue not conforming to a ductal distribution, geographic |
| ❑ Multiple regions | Enhancement in at least two large volumes of tissue not conforming to a ductal distribution, multiple geographic areas, patchy areas of enhancement |
| ❑ Diffuse | Enhancement distributed uniformly throughout the breast |

**Non-Mass Like Enhancement**

| **Internal Enhancement** *(select one)* | **Description** |
|---|---|
| ❑ Homogeneous | Confluent uniform enhancement |
| ❑ Heterogeneous | Nonuniform enhancement in a random pattern |
| ❑ Stippled, punctate | Punctuate, similar appearing enhancing foci, sand-like or dot-like |
| ❑ Clumped | Cobblestone-like enhancement, with occasional confluent areas |
| ❑ Reticular, dendritic | Enhancement with finger like projections extending toward nipple, especially seen on axial or sagittal images, in women with partly fatty–involuted breasts |

**D. Symmetric or Asymmetric (bilateral scans only)**

| **Symmetric or Asymmetric** *(select one)* | **Description** |
|---|---|
| ❑ Symmetric | Mirror-image enhancement |
| ❑ Asymmetric | More in one breast than in the other |

**E. Other Findings** *(select all that apply)*

| | |
|---|---|
| ❑ None apply | ❑ Edema |
| ❑ Nipple retraction | ❑ Lymphadenopathy |
| ❑ Nipple invasion | ❑ Pectoralis muscle invasion |
| ❑ Pre-contrast high ductal signal | ❑ Chest wall invasion |
| ❑ Skin thickening (focal) | ❑ Hematoma/blood |
| ❑ Skin thickening (diffuse) | ❑ Abnormal signal void |
| ❑ Skin invasion | ❑ Cysts |

## BI-RADS® – MRI

**F. Kinetic Curve Assessment**

| **Kinetic Curve Assessment** *(select one)* | **Description** |
|---|---|
| ❑ Initial rise | Slow, medium, rapid |
| ❑ Delayed phase | Persistent, plateau, washout |

**G. Assessment Category**

| **Assessment Category** *(select one)* | **Description** |
|---|---|
| ❑ **Category 0** – Incomplete: Need additional imaging evaluation | Finding for which additional evaluation is needed |
| **Final Assessment** | |
| ❑ **Category 1** – Negative | No abnormal enhancement, no lesion found (routine follow-up) |
| ❑ **Category 2** – Benign finding | Benign, no malignant features; i.e., cyst, (routine follow-up) |
| ❑ **Category 3** – Probably benign finding | Probably benign finding (short interval follow-up) |
| ❑ **Category 4** – Suspicious abnormality | Low to moderate suspicion for malignancy (biopsy should be considered) |
| ❑ **Category 5** – Highly suggestive of malignancy | High probability of malignancy (appropriate action should be taken) |
| ❑ **Category 6** – Known cancer | Biopsy-proven malignancy diagnosis on the imaged finding prior to definitive therapy (appropriate action should be taken) |