

Functional Magnetic Resonance Imaging of Breast Cancer



Gabrielle Carmen Baxter

Department of Radiology
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Downing College

November 2020

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the Degree Committee for Clinical Medicine and Veterinary Medicine.

Gabrielle Carmen Baxter
November 2020

Abstract

Functional Magnetic Resonance Imaging of Breast Cancer

Gabrielle Carmen Baxter

This thesis examines the use of magnetic resonance imaging (MRI) techniques in the detection of breast cancer and the prediction of pathological complete response (pCR) to neoadjuvant chemotherapy (NACT).

This thesis compares the diagnostic performance of diffusion-weighted imaging (DWI) models in the breast using a systematic review and meta-analysis. Advanced diffusion models have been proposed that may improve the performance of standard DWI using the apparent diffusion coefficient (ADC) to discriminate between malignant and benign breast lesions. Pooling the results from 73 studies, comparable diagnostic accuracy is shown using the ADC and parameters from the intra-voxel incoherent motion (IVIM) and diffusion tensor imaging (DTI) models. This work highlights a lack of standardisation in DWI protocols and methodology. Conventional acquisition techniques used in DWI often suffer from image artefacts and low spatial resolution. A multi-shot DWI technique, multiplexed sensitivity encoding (MUSE), can improve the image quality of DWI. A MUSE protocol has been optimised through a series of phantom experiments and validated in 20 patients. Comparing MUSE to conventional DWI, statistically significant improvements are shown in distortion and blurring metrics and qualitative image quality metrics such as lesion conspicuity and diagnostic confidence, increasing the clinical utility of DWI.

This thesis investigates the use of dynamic contrast-enhanced MRI (DCE-MRI) in the detection of breast cancer and the prediction of pCR. Abbreviated MRI (ABB-MRI) protocols have gained increasing attention for the detection of breast cancer, acquiring a shortened version of a full diagnostic protocol (FDP-MRI) in a fraction of the time, reducing the cost of the examination. The diagnostic performance of abbreviated and full diagnostic protocols is systematically compared using a meta-analysis. Pooling 13 studies, equivalent diagnostic accuracy is shown for ABB-MRI in cohorts enriched with cancers, and lower but not significantly different diagnostic performance is shown in screening cohorts.

Higher order imaging features derived from pre-treatment DCE-MRI could be used to predict pCR and inform decisions regarding targeted treatment, avoiding unnecessary toxicity. Using data from 152 patients undergoing NACT, radiomics features are extracted from baseline DCE-MRI and machine learning models trained to predict pCR with moderate accuracy. The stability of feature selection using logistic regression classification is demonstrated and a comparison of models trained using features from different time points in the dynamic series demonstrates that a full dynamic series enables the most accurate prediction of pCR.

Acknowledgements

I would first like to thank my supervisor Professor Fiona Gilbert for providing me with the opportunity to do this PhD. I am extremely lucky to have been supervised by such a celebrated and highly respected academic and expert in the field. This work would be nothing without your clinical insight and wealth of experience. Your encouragement and support has given me the confidence to pursue a career in research and for that I am eternally grateful.

I am extremely grateful to Dr Andrew Patterson for his supervision throughout this project and for teaching me how to operate an MR scanner.

I would like to thank Dr Martin Graves for his MR expertise, technical support, excellent proofreading and all of his contributions to this work. I am extremely grateful to Dr Ramona Woitek for all of her clinical insights and help with image analysis. I would like to acknowledge Dr Roie Manavaki, Dr Joshua Kaggie, Dr Mary McLean, Dr Iris Allabeju, Dr Julia Carmona-Bozo, Dr Frank Riemer, Dr James McKay, Dr Andrew Priest, Dr Leonardo Rundo and Dr Yuan Huang for all their help in the various projects making up this PhD.

I would like to thank the everyone at MRIS, particularly Ilse Patterson and Amy Schiller and the team of breast radiographers for carrying out patient scans (and for hours of entertaining conversation on the long days scanning at the Rosie!). I would also like to thank Simon Desonie from GE for technical support.

I am hugely indebted to the team of research nurses: Amy Frary, Johanna Field-Rayner, Sarah Hilborne and Jackie Mason for their help recruiting patients and for getting me set up to recruit patients (even if I did only manage one during my PhD!). To the departmental administrative staff, especially Candice Anderson and Sarah Perkins, who are not only extremely helpful and patient but also wonderfully kind. It is not an overstatement to say this PhD would not have happened without all of you- you keep the world turning. Thank you to everyone else on Levels 4 and 5 for creating such a lovely work environment - and also for the endless supply of cake and biscuits.

I would like to thank GE Healthcare for funding this PhD. I would also like to thank Downing College for awarding me a travel grant and the Oon Khye Beng Ch'hia Tsio grant to support this PhD.

Just like the Beatles, I get by with a little help from my friends.

A special thank you to Dimitri Kessler, Dr Sarah Hickman, and Zakaria Meddings, my departmental brothers in arms and the best people to complain to and eat snacks with on days where it seems like nothing is going right.

To Ami and Milena, the BFF crew, I wouldn't be where I am today without either of you. Thank you to Archie, Lowell, Sam, and Ross, who saw me through my salad days and 4 years of undergrad. To my huge, outrageous, and extremely dysfunctional Manchester uni family, I'm sorry for betraying you all and moving down South, I miss you all endlessly. Thank you to the colourful cast of characters I've had the pleasure of meeting at Downing, how lucky I am to have met friends for life in my 3 short years here- and to Dom, who set an example for all of us on how to make the most of the few years we have on this earth. A big thank you to Tom Bennett, in the hope that this finally puts a smile on his face. To the residents of 30 Lensfield road, it's been an honour to be your grandma Gaby- you guys keep me young. To Beth, my twisted sister, I am incredibly lucky to have you in my life.

To my sister Erin, thank you for all the laughs, vibe checks, and for being my greatest ally in the craziness that is the Baxter family. A special shout out to Milo, the best four-legged thesis writing companion you could ask for.

To my wonderful parents- who once upon a time were insistent that I get a "real job" like a "normal adult" but now will never hesitate to tell friends, colleagues, the Uber driver, the Sainsburys checkout lady, a fellow dog walker at the park, or complete strangers that you are parents of a Cambridge PhD student. Anything I have ever achieved in my life is a testament to the endless love and support you have given me. I can only hope that I've made you proud.

To Nick, my partner in crime, my biggest cheerleader, and my best friend, there are no words to express how grateful I am to have had you along with me for this crazy ride. You celebrate all of my successes and take all of my neuroses in your stride. You are patient and you are kind and I am a better version of myself because of you.

Finally, I would like to thank all the women who participated in our imaging studies. Knowing that this PhD may have some impact, however small, on the lives of women with breast cancer really does make it easier to get out of bed in the morning and makes all the bad days worth it. There is no more praise I could heap onto the NHS than has already been done this year, but to work alongside people who do lifesaving work every day is to truly stand amongst superheroes. To all the brave women out there fighting for their lives - this work is for you.

Publications arising from this thesis

Baxter, G.C., Selamoglu, A., Mackay, J.W., Bond, S., Gray, E., Gilbert, F.J., A meta-analysis comparing the diagnostic performance of abbreviated MRI (ABB-MRI) and a full diagnostic protocol (FDP-MRI) in breast cancer, *Clinical Radiology*, (2020)

Baxter, G.C., Patterson, A.J., Woitek, R., Allajbeu, I., Graves, M.J., Gilbert, F.J., Improving the image quality of diffusion-weighted imaging (DWI) in breast cancer: comparison of multi-shot DWI using multiplexed sensitivity encoding (MUSE) to conventional single-shot echo-planar imaging DWI, *The British Journal of Radiology*, (2020)

Baxter, G.C., Graves, M.J., Gilbert, F.J., Patterson, A.J., A meta-analysis of the diagnostic performance of diffusion MRI for breast lesion characterization, *Radiology*, 291(3):632-641, (2019)

Publications arising from work unrelated to this thesis

Carmona-Bozo, J.C., Manavaki, R., Woitek, R., Torheim, T., **Baxter, G.C.**, Caracò, C., Provenzano, E., Graves, M.J., Fryer, T.D., Patterson, A.J., Gilbert, F.J., Hypoxia and perfusion in breast cancer: a study using simultaneous PET/MR imaging, *European Radiology*, (2020)

Weir-McCall, J.R., Joyce, S., Clegg, A., MacKay, J.W., **Baxter, G.C.**, Dendl, L., Rintoul, R.C., Qureshi, N.R., Miles, K., Gilbert, F.J., Dynamic contrast-enhanced computer tomography for the diagnosis of solitary pulmonary nodules: a systematic review and meta-analysis, *European Radiology*, **30**, 3310–3323, (2020)

Presentations

Oral

Predicting complete pathological response (pCR) to neoadjuvant chemotherapy in breast cancer: machine learning-based comparison of radiomics features derived from pre-treatment DCE-MRI, ISMRM Virtual Conference, August 2020

Predicting complete pathological response (pCR) to neoadjuvant chemotherapy in breast cancer: features from DCE-MRI, CRUK Cancer Imaging Workshop, Cambridge, UK, February 2020

A comparative study of single-shot and multi-shot diffusion-weighted imaging (DWI) in breast cancer, CRUK Graduate Symposium, Cambridge, UK, July 2019

Poster

A meta-analysis comparing the diagnostic performance of abbreviated MRI (ABB-MRI) and a full diagnostic protocol (FDP-MRI) in breast cancer, ISMRM Virtual Conference, August 2020

Quantitative Bilateral Sodium and Proton Breast Imaging at 3T, ISMRM Virtual Conference, August 2020

Diffusion-weighted imaging using multiplexed sensitivity encoding (MUSE) in breast cancer: comparison with single-shot echo-planar imaging, ISMRM Virtual Conference, August 2020

A comparative study of multi-shot diffusion-weighted imaging (DWI) using multiplexed sensitivity encoding (MUSE) and conventional single-shot echo-planar imaging (EPI) DWI in breast cancer, ISMRM, Montreal, Canada, May 2019

A meta-analysis comparing the diagnostic performance of Diffusion-Weighted Imaging (DWI), Diffusion Tensor Imaging (DTI), and Intra-voxel Incoherent Motion (IVIM) in breast cancer, ISMRM, Montreal, Canada, May 2019

Nomenclature

Abbreviations/Acronyms

ABB-MRI	Abbreviated MRI
ADC	Apparent diffusion coefficient
AUC	Area under the curve
CI	Confidence intervals
DCE	Dynamic contrast-enhanced
DKI	Diffusion kurtosis imaging
DTI	Diffusion tensor imaging
DWI	Diffusion-Weighted imaging
ER	Estrogen receptor
FDP-MRI	Full diagnostic protocol MRI
HER2	Human epidermal growth factor receptor 2
IVIM	Intra-voxel incoherent motion
LASSO	Least absolute shrinkage and selection operator
MI	Mattes mutual information
MRI	Magnetic resonance imaging
MUSE	Multiplexed sensitivity encoding
NACT	Neoadjuvant chemotherapy

NMR	Nuclear magnetic resonance
pCR	Pathological complete response
PR	Progesterone receptor
RF	Radiofrequency
ROC	Receiver operating characteristics
SD	Standard deviation
SNR	Signal-to-noise ratio
ss-EPI	Single-shot echo-planar imaging
T1W	T ₁ -weighted
T2W	T ₂ -weighted
TRICKS	Time resolved imaging contrast kinetics
VIBRANT	Volume image based assessment

Mathematical symbols

α	Flip angle
γ	Gyromagnetic ratio
\hbar	Reduced Planck's constant
μ	Magnetic dipole moment
ω_0	Larmor frequency
ϕ	Phase
ρ	Proton density
b	b-value
B ₀	Static magnetic field
M ₀	Net magnetisation

Table of contents

Publications arising from this thesis	ix
Presentations	xi
Nomenclature	xiii
List of figures	xix
List of tables	xxi
1 Introduction	1
1.1 Breast cancer and the role of imaging	1
1.2 Imaging the tumour microenvironment	2
1.2.1 Breast cancer pathology	2
1.2.2 Imaging vascularity and angiogenesis: Qualitative and quantitative assessment of DCE-MRI	5
1.2.3 Imaging tumour cellularity and microstructure: Diffusion-weighted imaging	7
1.3 MRI screening	10
1.3.1 Abbreviated MRI	10
1.3.2 Non-contrast-enhanced screening	11
1.4 Neoadjuvant chemotherapy for breast cancer	12
1.4.1 Therapeutic targets for chemotherapy	12
1.4.2 Pathological complete response to neoadjuvant chemotherapy . . .	13
1.4.3 Monitoring response to neoadjuvant chemotherapy	14
1.5 Thesis aims and outline	17
2 Magnetic Resonance Imaging	21
2.1 Nuclear magnetic resonance	21

2.1.1	Nuclear magnetism	21
2.1.2	Macroscopic magnetisation	24
2.1.3	NMR signal generation	25
2.1.4	Mechanisms of relaxation	26
2.2	Magnetic resonance imaging: hardware and image formation	30
2.2.1	Static magnetic field	31
2.2.2	Radiofrequency (RF) coils: NMR signal generation and detection	31
2.2.3	Gradient coils: spatial localisation	32
2.2.4	Image formation	37
2.3	Pulse sequences and acquisition techniques	38
2.3.1	Spin echo	39
2.3.2	Gradient echo	40
2.3.3	Generating image contrast	42
2.3.4	Signal and noise	44
2.3.5	Fat suppression	44
2.3.6	Parallel imaging	47
2.4	Dynamic contrast-enhanced MRI	48
2.4.1	Temporal acceleration	49
2.5	Diffusion-weighted imaging	51
2.5.1	Diffusion-weighted preparation	51
2.5.2	DWI artefacts	53
2.5.3	Multi-shot imaging	54
2.5.4	Advanced diffusion models	55
3	A meta-analysis of the diagnostic performance of diffusion MRI for breast lesion characterisation	59
3.1	Introduction	59
3.2	Methods	60
3.2.1	Literature search	60
3.2.2	Study selection	60
3.2.3	Data extraction	61
3.2.4	Data quality assessment	61
3.2.5	Statistical analysis	61
3.3	Results	62
3.3.1	Study selection and data extraction	62
3.3.2	Data quality assessment	64
3.3.3	Statistical analysis	64

3.4	Discussion	69
4	Improving the image quality of diffusion-weighted imaging (DWI) in breast cancer	71
4.1	Introduction	71
4.2	Methods	72
4.2.1	Phantom study design	72
4.2.2	Patient study design	74
4.2.3	Image analysis	75
4.3	Results	76
4.3.1	Phantom results	76
4.3.2	Patient results	77
4.3.3	Qualitative results	80
4.3.4	Quantitative results	80
4.3.5	Comparison of ADC values: ss-EPI versus MUSE	80
4.4	Discussion	82
5	A meta-analysis comparing the diagnostic performance of abbreviated MRI and a full diagnostic protocol in breast cancer	85
5.1	Introduction	85
5.2	Methods	86
5.2.1	Literature search	87
5.2.2	Study selection	87
5.2.3	Data extraction	87
5.2.4	Data quality assessment	88
5.2.5	Statistical analysis	88
5.3	Results	88
5.3.1	Study selection and data extraction	88
5.3.2	Data quality assessment	90
5.3.3	Statistical analysis	94
5.4	Discussion	97
6	Predicting pathological complete response to neoadjuvant chemotherapy in breast cancer: machine learning-based analysis of radiomics features derived from pre-treatment DCE-MRI	101
6.1	Introduction	101
6.2	Methods	103

6.3	Results	108
6.4	Discussion	117
7	Future work and conclusions	123
7.1	Contributions to knowledge	123
7.2	Future work	124
7.3	Conclusions	127
	References	129

List of figures

1.1	Diagnostic indices of imaging methods	2
1.2	Breast cancer anatomy	3
1.3	Time-signal intensity curves	6
1.4	Toft's model	7
1.5	Restricted diffusion	8
2.1	The interaction of a nucleus with a magnetic field	23
2.2	The two possible spin states of nuclei	24
2.3	Net magnetisation	25
2.4	90° RF pulse	26
2.5	Simulated free induction decay (FID).	27
2.6	Calculated T ₁ and T ₂ relaxation curves	29
2.7	Schematic of an MRI scanner	30
2.8	Slice selection	33
2.9	Frequency encoding gradient	34
2.10	Spatial localisation using frequency encoding	35
2.11	Phase encoding	36
2.12	Repeating the phase encoding	36
2.13	<i>k</i> -space diagram	37
2.14	MR image reconstruction	39
2.15	Spin echo	40
2.16	Fast spin echo	41
2.17	Gradient echo	42
2.18	Short TI inversion recovery	45
2.19	Parallel acceleration	47
2.20	3D Gradient echo	50
2.21	TRICKS	50
2.22	Diffusion-weighted imaging pulse sequence	52

2.23	<i>k</i> -space trajectories used for diffusion-weighted imaging	55
2.24	Non-Gaussian diffusion	56
3.1	PRISMA flow diagram	63
3.2	QUADAS-2 results	64
3.3	Funnel plots	65
3.4	Forest plots	67
3.5	Summary receiver operating characteristics curves	68
4.1	The breast phantom	73
4.2	Images of the breast phantom using MUSE	76
4.3	Phantom metrics	77
4.4	Comparisons of image quality for ss-EPI and MUSE	78
4.5	Quantitative comparisons of image quality for ss-EPI and MUSE	81
4.6	Comparisons of ADC and nADC for ss-EPI and MUSE	82
5.1	PRISMA flow diagram	89
5.2	QUADAS-2 results	90
5.3	Forest plots	96
5.4	Summary receiver operating characteristics curves	97
6.1	Radiomics analysis workflow.	105
6.2	Summary of the models trained using different combinations of features and post-contrast time points.	110
6.3	Performance of models trained using different feature sets	111
6.4	Correlation matrix	113
6.5	Visualisation of shape and histogram features commonly selected by the models	114
6.6	Performance of models trained using different time points	116

List of tables

1.1	Features of molecular subtypes of cancer	5
1.2	RECIST criteria for tumour response to treatment	14
2.1	Nuclei used in human MRI studies	22
2.2	In vivo relaxation times in the breast	30
2.3	Generating image contrast	43
3.1	Combined approaches using DTI and IVIM parameters	63
3.2	Results of pooled estimates and heterogeneity measures	66
3.3	Results of sub-analysis comparing studies using the ADC	68
4.1	Scan parameters	74
4.2	ADC values measured for varying PVP concentrations	77
4.3	Patient and lesion characteristics.	79
4.4	Qualitative comparisons	80
4.5	Quantitative comparisons	81
5.1	Characteristics of included screening studies	91
5.2	Characteristics of included enriched cohort studies	92
5.3	Technical details of included studies	93
5.4	Results of pooled analysis	95
6.1	Features sets used to train logistic regression models	106
6.2	Patient and lesion characteristics	109
6.3	Performance of models trained using different feature sets	112
6.4	Performance of models trained using different time points	117
6.5	Performance of models trained using combinations of time points	118

Chapter 1

Introduction

1.1 Breast cancer and the role of imaging

Breast cancer is the second most commonly diagnosed cancer in the world, accounting for 11% of all cancers, and the most commonly diagnosed cancer in women worldwide [1]. In the UK, over 55,000 women are diagnosed with breast cancer every year [2]. However, survival rates are improving with developments in non-surgical treatment, such as neoadjuvant chemotherapy and radiotherapy, and with early detection and diagnosis through improved imaging and screening [3].

Mammography is the most commonly used modality in breast cancer screening. While mammography has significantly reduced mortality, mammographic sensitivity is affected by breast density as well as tumour subtype [4]. For women with dense breast tissue there is a reduction in sensitivity from 85% overall to 47.8–64.4% [5, 6]. Ultrasound is often used as a supplemental screening tool to detect cancers in dense breasts, as well as to characterise abnormalities on mammograms, measure tumour size, assess nodal metastases or to guide needle biopsies of tumours.

However, magnetic resonance imaging (MRI) is the most sensitive technique for the detection of breast cancer (Figure 1.1) and has an established role in the assessment of response to chemotherapy and overall breast cancer management. MRI is often used as an adjunct imaging method to investigate equivocal findings on mammography and ultrasound and is used world wide in the screening of high-risk women (including those with elevated familial risk of breast cancer or BRCA1 and BRCA2 carriers) [7, 8].

Dynamic contrast-enhanced MRI (DCE-MRI) achieves a high sensitivity of 93% in the detection of breast cancer [10]. DCE-MRI consists of a series of three-dimensional image volumes acquired before, during and after the intravenous administration of a gadolinium-based contrast agent. As well as providing high resolution anatomical and morphological

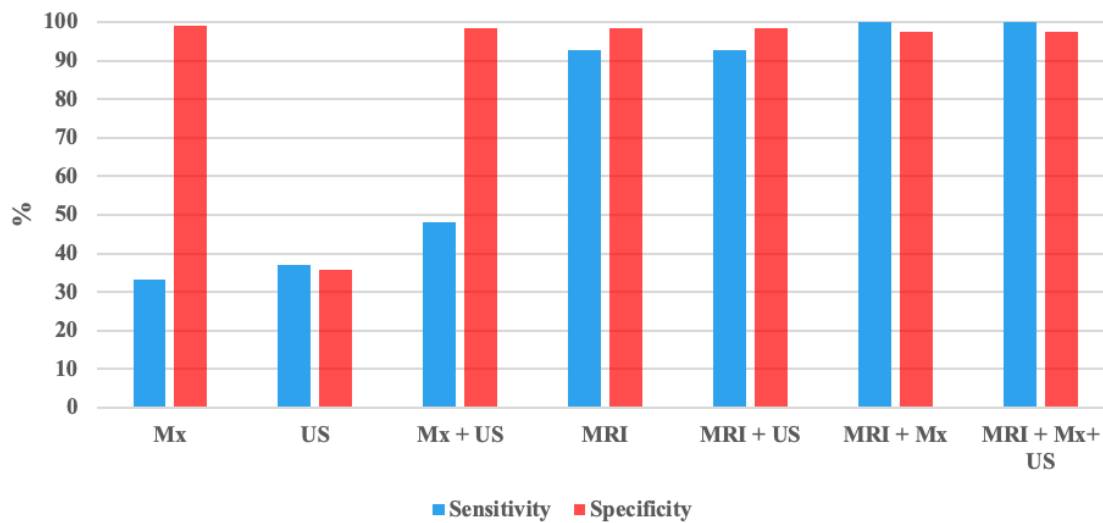


Figure. 1.1 Diagnostic indices of imaging methods used alone or in combination for women at an elevated risk of breast cancer [9]. Mx, mammography; US, ultrasound; MRI, magnetic resonance imaging

information, it can also provide functional information regarding tumour perfusion and vascularity. Diffusion-weighted imaging (DWI) is a non-contrast MRI technique that is sensitive to the random Brownian motion of water molecules in the extracellular space. Measurement of the apparent diffusion coefficient (ADC) using DWI can quantify the restriction of molecular diffusion in biological tissues, providing complementary functional information that reflects tissue cellularity and the integrity of cellular membranes and structures. DWI is often used as an adjunct to DCE-MRI in the diagnosis of breast cancer to reduce false-positive results, achieving a combined sensitivity and specificity of 91.6% and 85.5%, respectively [10].

1.2 Imaging the tumour microenvironment

1.2.1 Breast cancer pathology

Breast cancer is a highly heterogeneous disease. It can be characterised by its histopathological subtype, histologic grade, and the expression of proteins and genes, which can be used to select appropriate treatment, such as use or choice of chemotherapy and extent of surgery. Pathologic and molecular characteristics of breast tumours can also be used as prognostic factors to estimate overall disease outcomes (such as disease-free survival and overall survival

in the absence of therapy), or predictive factors to estimate the likelihood of response to specific treatment (such as the PREDICT score [11]).

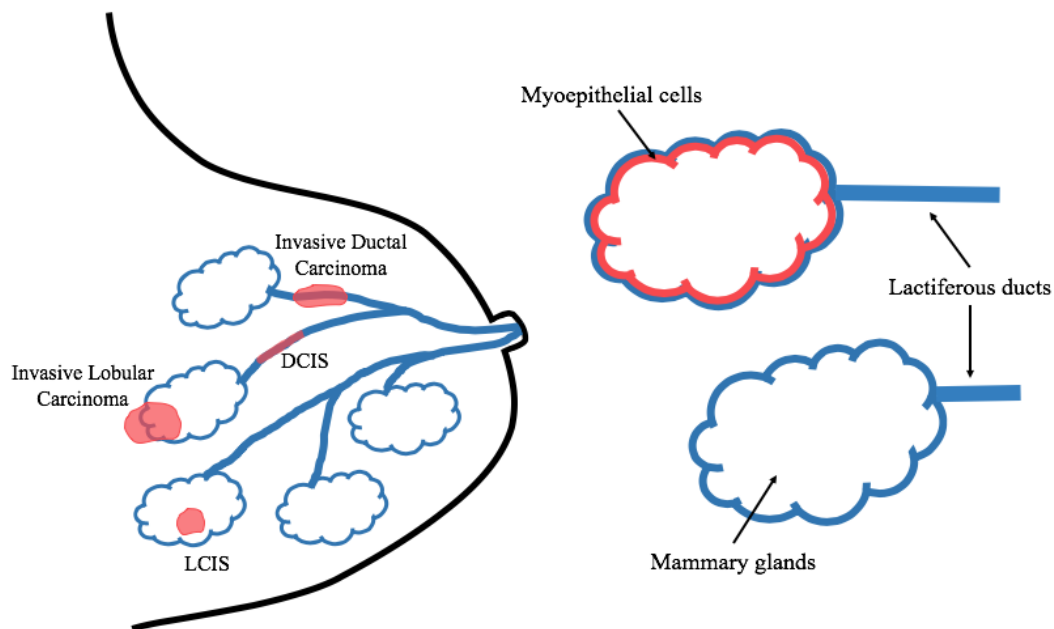


Figure. 1.2 Breast cancer anatomy

The majority of invasive breast cancers are adenocarcinomas, derived from the epithelial tissue (the tissue forming the outer layer) of breast ducts or lobules, shown in Figure 1.2. In situ, or pre-invasive, carcinomas are those in which the cancer cells have not yet infiltrated tissue surrounding the ducts or lobules, such as ductal carcinoma in situ (DCIS) and lobular neoplasia (including lobular carcinoma in situ, or LCIS) which originate in and are confined to the breast ducts and lobules. Invasive carcinomas are those in which the cancer cells have undergone additional mutations to spread out from the basement membrane and myoepithelial cells of the ducts and lobules into surrounding tissue and have the ability to metastasise. Microscopy is used to determine the histopathologic classification from biopsied specimens.

The most common invasive breast carcinoma is invasive ductal carcinoma (IDC), or invasive carcinoma of no special type (NST), accounting for around 75% of all breast cancers. Invasive lobular carcinoma (ILC) accounts for 10-15% of all breast cancers and originates from the lobular epithelium. ILCs have different clinical and prognostic factors to other cancers and generally have a low histologic grade and low to intermediate cell proliferation [12, 13]. There are a number of less common subtypes such as tubular, medullary, mucinous, and papillary carcinoma. While rare subtypes do correlate with prognosis, histopathological subtype is generally not considered in the clinical management of breast cancer [14].

The histologic grade of breast cancer, a measure of how different tumour cells are to normal breast tissue, is strongly correlated with prognosis [15]. A numerical scoring system with an overall grade of 1-3, known as the Nottingham Grading System, was developed by Elston and Ellis; cells in grade 1 tumours tend to be small and divide slowly whereas cells in grade 3 cancers are larger and rapidly dividing; patients with grade 1 tumours have significantly better survival than those with grade 2 or 3 tumours [16]. Grade is determined by the evaluation of three morphological features: tubule formation (how different tumour tissue looks compared to normal breast tissue), nuclear pleomorphism (how different tumour cell nuclei are compared to normal nuclei), and mitotic count (a measure of cellular proliferation).

Hormone receptor status provides important prognostic information and is critical in the selection of appropriate targeted treatment. The presence of estrogen receptors (ER), progesterone receptors (PR), and human epidermal growth factor receptor 2 (HER2) in breast tumours are determined by immunohistochemistry or multigene assay. Estrogen receptors are over-expressed (ER+) in 70% of breast cancers [17]. Cumulative exposure of the breast epithelium to estrogen has been shown to be a risk factor associated with breast cancer [18]. It has been suggested that the binding of estrogen to estrogen receptors stimulates the proliferation of mammary cells which elevates the risk of replication errors and produces genotoxic by-products that damage DNA [19]. The role of PR status in cancer risk and tumour progression is less clear, though it has been shown that cancers that are progesterone receptor negative (PR-) are associated with a significant reduction in disease free survival and overall survival for ER+ breast cancer [20]. HER2, which regulates cell growth and proliferation, is over-expressed (HER2+) in 20-30% of breast cancers [17]. HER2 positive cancers have historically been associated with more aggressive disease and poor prognosis [21], however the development of HER2-targeted therapy has improved outcomes [22].

It has been shown that there are five clinically relevant molecular subtypes based on gene expression profiling, with distinct clinical outcomes [23, 24] (Table 1.1). Further work by Curtis et al. [25] has shown there are potentially 10 different subtypes of breast cancer. These are determined by ER, PR and HER2 status, as well as Ki-67 expression, a marker of cellular proliferation. Four of the five subtypes can be mapped to an immunohistochemical subtype: there are two ER positive subgroups (Luminal A and B, with the Luminal B subtype usually having a higher grade) and two predominantly ER negative subgroups (HER2-enriched with overexpression of HER2 and related genes and basal-like or triple negative characterised by high expression of genes associated with myoepithelial/basal cells). Normal-like cancers share an immunohistochemical subtype with Luminal A cancers and are characterised by normal breast tissue profiling [24].

Table 1.1 Features of molecular subtypes of cancer [26].

Subtype	ER	PR	HER2	Ki-67 expression	Frequency (%)
Luminal A	+	+	-	Low	29
Luminal B	+	+	+/-	High	21
Triple-negative/Basal	-	-	-	High	16
HER2-enriched	-	-	+	Moderate-high	17
Normal-like	+	+	-	Low	6

Different molecular subtypes of breast cancer are associated with different outcomes. Luminal A patients have been found to generally have the best prognosis [27]. HER2-enriched and triple-negative cancers are associated with a poor prognosis compared to Luminal and normal-like cancers [23], with triple-negative cancers representing a large proportion of breast cancer-related deaths due to their aggressive phenotypic characteristics and lack of targeted therapies [28].

1.2.2 Imaging vascularity and angiogenesis: Qualitative and quantitative assessment of DCE-MRI

Malignant tumours undergo angiogenesis and develop complicated capillary networks to meet the increased demand for oxygen and nutrients due to increased cell proliferation [29]. However, tumour vasculature develops differently to normal tissue, where blood is supplied to tissues from an ordered and efficient mature vascular network. In a tumour, the population of cells grows aggressively, leading to the development of a disorganised network of immature, tortuous and hyperpermeable blood vessels [30]. Grubstein et al. studied the blood supply to a tumour site using DCE-MRI, finding that breast lesions had an altered general vascular supply, a prominent feeding vessel and increased regional vascularity [31]. Angiogenesis has a direct impact on the uptake of contrast agent and therefore the appearance of lesions on DCE-MRI due to the density and distribution of blood vessels [32], leading to the widespread use of qualitative and quantitative DCE-derived features in the diagnosis of malignant or benign breast cancer. The introduction of the Breast Imaging Reporting and Data System (BI-RADS) descriptors [33] by the American College of Radiology standardised the terminology used to report breast MRI, including morphological features (such as shape and margins) and internal enhancement characteristics for mass and non-mass lesions, assessment of background parenchymal enhancement and amount of fibroglandular tissue, kinetic curve assessment, and other associated features.

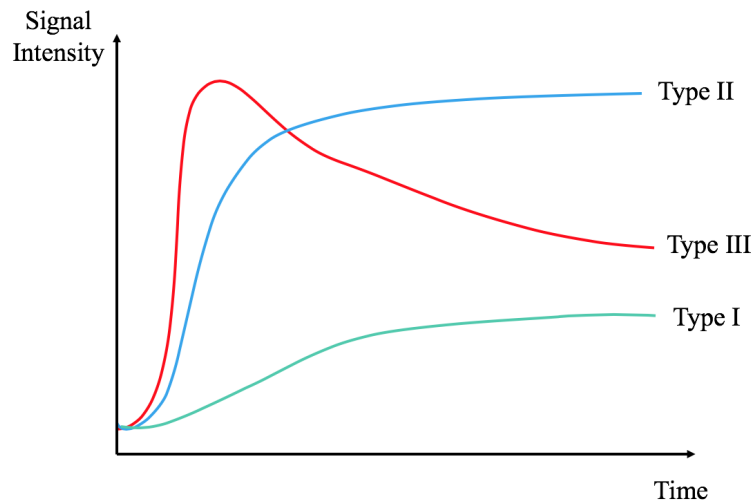


Figure. 1.3 Time-signal intensity curves classified as steady (type I), plateau (type II) or washout (type III) [34].

Acquiring a series of MR images over the time course of an injection of contrast agent allows a kinetic curve of tissue enhancement to be produced, visualising the dynamic flow of contrast through tissue. Kuhl et al. developed a classification scheme based on patterns of contrast uptake that is widely used to differentiate between malignant and benign lesions [34] (Figure. 1.3)

Type I curves show a steady gradual enhancement over time, type II curves show a rapid initial enhancement with a plateau, and type III curves show a rapid initial enhancement with washout. Type I curves were observed in 83% of benign cases and type II and III curves were observed in 91% of malignant cases (34% type II and 57% type III) [34]. In malignant lesions, this corresponds to an increased but leaky vasculature where the contrast agent pools then escapes. In benign lesions with a more normal vasculature, there is a gradual retention of contrast agent until saturation. These enhancement curves can be used to calculate a number of heuristic parameters such as signal enhancement ratio, time to peak enhancement, wash-in and wash-out gradient, and initial area under the time signal curve (iAUC).

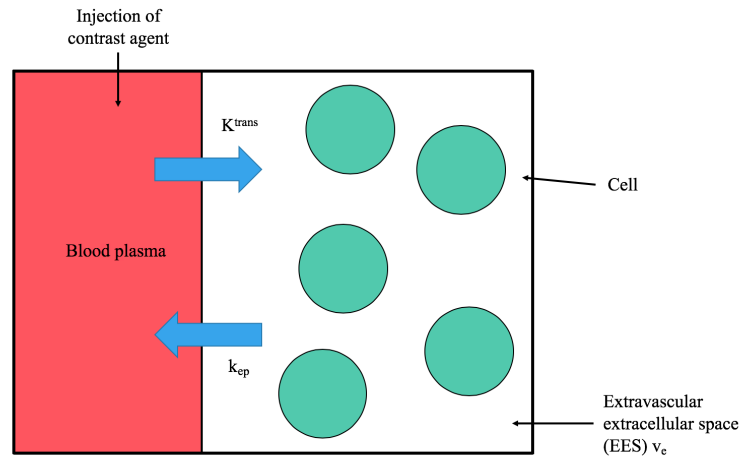


Figure. 1.4 The two compartment Toft's model. Pharmacokinetic modelling can be used to fit parameters to time-intensity curves to reflect perfusion and microvessel permeability.

Pharmacokinetic modelling is also used to characterise the circulation of contrast agent through the vasculature. A simple two-compartment model of the blood plasma and the extravascular extracellular space (EES) is used to derive parameters to describe aspects of tumour physiology such as perfusion, intravascular and extravascular volume fractions, microvessel wall permeability, and microvessel density. The most commonly used parameters are described by the Toft's model [35], including the transfer coefficient between the intravascular and extravascular spaces K^{trans} (min^{-1}), the proportion of extravascular extracellular space v_e , and the flux rate constant between the extravascular extracellular space and the plasma k_{ep} (min^{-1}), as shown in Figure. 1.4.

1.2.3 Imaging tumour cellularity and microstructure: Diffusion-weighted imaging

Unlike homogeneous materials, the diffusion of water in biological tissues is not isotropic and is restricted due to physical barriers and microstructures. Fluid viscosity, permeability between intra- and extracellular membranes, and active transport and flow impede or enhance the diffusion of water in tissue and ordered structures introduce a directional dependence to diffusion [36]. Blood flow and bulk tissue motion from respiration and cardiac pulsation also affect the measurement of water diffusion in tissue. As such, the measurement of diffusion in tissue using DWI was termed the 'apparent' diffusion coefficient (ADC)(mm^2/s). The uncontrolled proliferation of tumour cells results in a high tumour cellularity, a more restricted

extracellular environment, and a loss of normal tissue architecture, which has been shown to be an early indicator of abnormality [37].

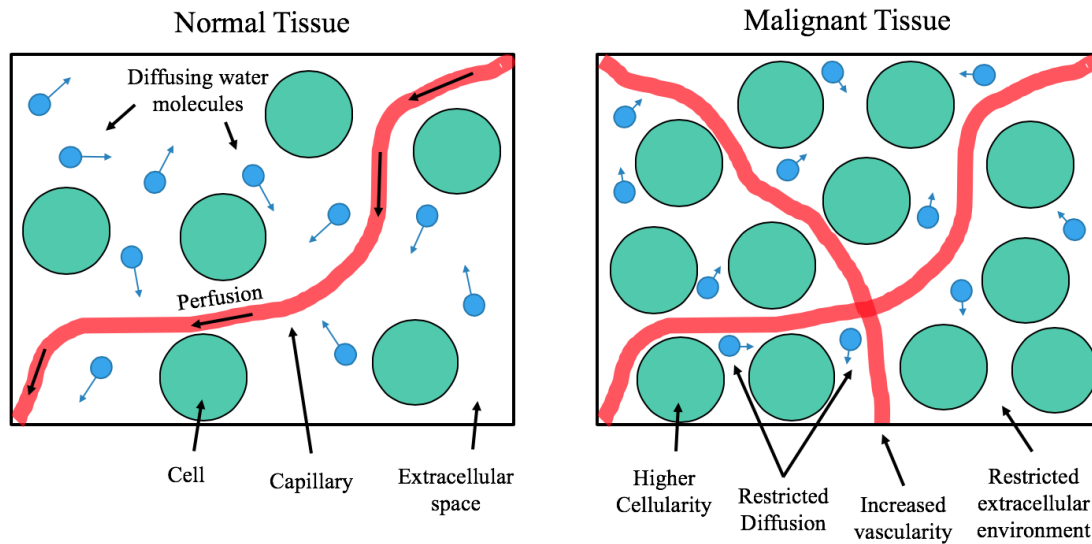


Figure. 1.5 Restricted diffusion. The mobility of diffusing water molecules in malignant breast tumours is decreased due to higher cellularity and a more restricted extracellular environment.

Significantly lower ADC values are measured in malignant breast tumours compared to benign lesions and fibroglandular tissue [38], indicating more restricted diffusion, as shown in Figure 1.5. Initial studies in the breast compared the ADC as measured using DWI to histological examination of tissues, finding that the ADC was inversely correlated with cellularity [39, 40]. Hatakenaka et al. suggested that the lower extracellular water content and increased barrier structures resulted in more restricted diffusion [40]. Choi et al. found lower ADC values to be significantly correlated with higher cell proliferation measured using the Ki-67 index [41]. Lower ADC has also been shown to be correlated with tumour aggressiveness [42] and higher histological grade, larger tumour size and presence of positive axillary lymph nodes [43].

DWI is increasingly used in breast lesion diagnosis and characterisation. The ADC achieves a high diagnostic performance in the differential diagnosis of malignant and benign breast lesions [38]. Marini et al. demonstrated that the ADC of invasive ductal carcinoma was low compared to other histopathological subtypes due to more densely packed tumour cells [44]. A number of false negative and false positive findings on DWI can be attributed to differences in underlying tumour histology. Woodhams et al. showed that the ADC of mucinous carcinoma was significantly higher than other malignant tumours due to the low

cellularity relative to the abundant mucin, as well as high extracellular water content and a mucin pool [45]. Higher ADC values are also measured for papillary carcinomas due to the distribution of tumour cell batches within stromal spaces, allowing for more free diffusion in the interstitium [46]. Central tumour necrosis results in a higher ADC due to a decrease in overall tumour cellularity. Marini et al. found a statistically significant difference in ADC between the centre and edges of locally advanced cancers which corresponded a large necrotic centre and higher cellular density in the edges as seen on histopathology [44]. The most commonly reported false positive lesions on DWI are complicated cysts and fibroadenomas, likely due to high cellularity, fibrosis and chronic inflammatory elements [47, 48].

Advanced diffusion models

Other models of diffusion have been proposed that aim to probe aspects of diffusion behaviour in the complex tumour microenvironment that are not captured by the ADC model. Another consequence of the complicated capillary network developed through angiogenesis is the contribution of the perfusion of blood through the capillary network to the microscopic diffusion measured using DWI. The intra-voxel incoherent motion (IVIM) model of diffusion was proposed by Le Bihan et al. [49] which separates the contributions from tissue and the microvasculature into separate parameters as tissue diffusivity, D , and pseudo-diffusivity from perfusion, D^* , and measures the perfusion fraction, f . IVIM is increasingly investigated in the breast and has been able to distinguish between malignant and benign lesions with a high accuracy [50]. IVIM has also been used to identify molecular prognostic factors and predict treatment response [51].

Diffusion kurtosis imaging (DKI) was introduced by Jensen et al. to quantify the degree to which water diffusion in tissues is non-Gaussian [52]. DKI has been used in the characterisation of breast lesions as kurtosis has been shown to be higher in malignant lesions than in benign lesions and fibroglandular tissue [53, 54], suggesting higher microstructural complexity. Fibroadenoma and fibrocystic changes showed significant differences from normal tissue only in kurtosis [53]. The stretched exponential model, proposed by Bennett et al. [55] has also been used to model the non-Gaussian behaviour of diffusion and has been used in the differential diagnosis of breast tumours with a high sensitivity and specificity (81% and 82%, respectively) [56], as well as to assess response to neoadjuvant chemotherapy [57]. While the DKI and stretched exponential models have been shown to achieve a higher goodness-of-fit to data than IVIM and the ADC [56], these models are not associated with any physiological phenomena and the sources of non-Gaussian behaviour are unclear.

Tumour microstructures introduce a directional dependence on the restriction of diffusion. Diffusion tensor imaging (DTI) probes the anisotropy of diffusion by measuring components

of diffusion in many directions and resolves them into three orthogonal eigenvectors. The mean diffusivity (MD), the average of the diffusion eigenvalues, is analogous to the apparent diffusion coefficient and is similarly reduced in malignant lesions compared to benign, and has been used in the differential diagnosis of breast lesions[58–60]. Measures of anisotropy such as the fractional anisotropy have also been shown to be indicators of malignancy [59, 61, 62]. While DTI is most commonly used for tractography of white matter in the brain, it has been used to image the structural organisation of ductal and glandular trees in the breast [63, 64].

Challenges of DWI

While the use of DWI in the assessment of breast cancer has increased, in clinical practice DWI suffers from poor image quality due to the rapid acquisition using single-shot echo-planar imaging. Images are limited by spatial distortion and blurring, resulting in a low spatial resolution and the averaging of tumour volumes with adjacent breast tissue. Performance is significantly reduced for cancers less than 1cm in size [65, 66]. However, a number of alternative acquisition strategies have been investigated that attempt to overcome these limitations and improve the clinical utility of DWI [67]. The physics underlying DWI acquisition techniques will be discussed in Chapter 2.

1.3 MRI screening

Currently, MRI screening is only recommended for high-risk women (particularly those with a history of prior thoracic radiotherapy and strong familial risk of breast cancer, especially BRCA1/2 carriers) who are invited for annual examination from ages 25-50 (continuing over the age of 50 for women with dense breasts). However, the high cost of MRI, owing to the capital cost of equipment, long examination times, use of intravenous gadolinium-based contrast agent (GBCA), the need to check for renal function, and onsite costly medical staff to manage potential contrast reactions, as well as the increased time taken for the radiologist to review MRI images compared to mammography, limits the widespread use of MRI as a screening tool for moderate-risk women with dense breasts. It has been shown that MRI screening is currently only cost effective in very high-risk women [68, 69].

1.3.1 Abbreviated MRI

The use of abbreviated MRI (ABB-MRI) protocols for the detection of breast cancer has gained increasing attention. While a full diagnostic MRI protocol (FDP-MRI) includes a combination of unenhanced sequences (T_1 , T_2 , and diffusion-weighted imaging), as well

as contrast-enhanced sequences (DCE-MRI) acquired in a 30-minute examination time, an abbreviated protocol generally includes a non-contrast T_1 -weighted sequence and at least one contrast-enhanced T_1 -weighted sequence from which subtraction and 3D maximum-intensity projection (MIP) images can be generated. This substantially reduces the acquisition time as well as the radiologist interpretation time. Kuhl et al. reported the first prospective reader study of an abbreviated breast MRI protocol in a cohort of asymptomatic mild- to moderate-risk women achieving a sensitivity of 91%, reading MIP images generated from the first post-contrast sequence with an average reading time of 3 seconds [70]. Abbreviated protocols have since been increasingly investigated in varied patient populations including screening cohorts (with low, moderate, and high risk, as well as women with dense breasts), problem solving and preoperative staging patients, and screening cohorts enriched with cancers. Furthermore, the reported abbreviated protocols vary by institution, with studies investigating effect of the addition of a second post-contrast time point [71], T_2 -weighted imaging [72] or DWI [73] on diagnostic performance.

As abbreviated MRI is able to reduce the cost, examination time and interpretation time of MRI, this may enable the use of MRI as a supplemental imaging modality in those women with dense breasts, for risk-adaptive screening, or allow for the more widespread use of MRI screening for moderate-risk women. Investigating ABB-MRI in a cohort of mild- to moderate-risk women, Kuhl et al. found a cancer detection rate similar to that of a routine full diagnostic protocol in high-risk women (18.2 vs 17-22.1 per 1000) [70, 74, 75]. Given the reduction in mammographic sensitivity with increased breast density [5, 6], supplemental MRI screening for women with dense breasts has been investigated, with the DENSE trial of 40,373 women finding a reduced number of interval cancers detected than screening using mammography alone [76]. Abbreviated MRI could provide a lower cost solution for MRI screening in this population. In a screening cohort of 1444 women with dense breasts, Comstock et al. found a higher invasive cancer detection rate using ABB-MRI than digital breast tomosynthesis (DBT) [77]. A similar study by Weinstein et al. found a cancer detection rate of 24.7 per 1,000 using ABB-MRI in a cohort of 475 women with dense breasts and negative findings on DBT [78].

1.3.2 Non-contrast-enhanced screening

As well as cost, there are concerns about the repeated administration of gadolinium-based contrast agents (GBCAs) for contrast-enhanced screening, and the benefits and risks of screening MRI for high-risk women who are regularly examined over long-term screening periods must be considered. GBCAs can cause allergic or physiological reactions (such as nausea or headaches) in a small percentage of patients [79] and the use of contrast agent may

be an issue for patients with impaired renal function who are at risk for nephrogenic systemic fibrosis. Recently, studies have also noted the presence of gadolinium deposits in the brain and body with cumulative dose, though no clinical adverse side effects have yet been reported [80]. This is of particular interest for high-risk healthy women who undergo routine annual screening with DCE-MRI up to 40 times in their lifetime.

DWI could provide a non-contrast alternative to screening protocols using DCE-MRI. DWI has demonstrated a sensitivity and specificity comparable to contrast-enhanced MRI [50] and has been investigated as a supplemental imaging modality to exclude malignancy in women with suspicious mammograms [81]. A number of studies have investigated DWI used independently for screening or assessed the detection of cancers using non-contrast MRI with DWI through blinded reader studies, reporting sensitivities $> 85\%$ [48, 82–84] or a modest sensitivity (50–77%) and a high specificity $> 90\%$ [65, 66, 85]. DWI is also able to overcome some of the shortfalls of mammography. When compared to mammography, DWI performs better in the detection of smaller cancers $\leq 2\text{cm}$ [66], non-palpable cancers in asymptomatic women [85], and cancers in women with dense breasts where lesion visibility was shown to be significantly superior to mammography using DWI [86]. However, the performance of DWI is reduced for small cancers less than 1cm in size [65, 66]. For DWI to be clinically useful in a screening setting it must be able to detect and characterize all cancers, particularly small cancers. Detecting smaller cancers is crucial in reducing breast cancer mortality, as the chance of survival at 10 years improves from 60% for cancers detected at 2–5cm at screening to 85% when detected at 1–2cm [87].

1.4 Neoadjuvant chemotherapy for breast cancer

1.4.1 Therapeutic targets for chemotherapy

There are many approaches for the management of breast cancer including surgery, radiotherapy, endocrine treatment and chemotherapy. However, breast cancer is a heterogeneous disease, and treatment plans are chosen depending on a combination of factors such as tumour size, molecular subtype, axillary node status and existence of metastases. Neoadjuvant chemotherapy (NACT) is used in the management of breast cancer with a primary aim to reduce the size of a tumour before surgery. This can allow for less radical procedures, such as the option of breast conserving surgery for women who had originally required mastectomy. Axillary nodal response has also been shown to correlate with primary breast tumour response to neoadjuvant chemotherapy, reducing the need for axillary nodal dissection

[88]. Chemotherapy regimens are commonly given as a series of treatments, also known as cycles, with a break between treatments to allow the body to recover from side-effects.

Endocrine therapy is the primary therapy for ER and PR positive cancers. Hormones are modulated using drugs such as tamoxifen to reduce the effect of estrogen, and aromatase inhibitors can be used to lower the levels of estrogen in the body. Hormone receptor negative and HER2 positive breast cancers are treated with monoclonal antibodies targeted against HER2, such as trastuzumab and pertuzumab, which has improved outcomes and reduced morbidity [89]. Triple negative breast cancers are difficult to treat as drugs targeted at downregulating estrogen, progesterone and HER2 are ineffective [90]. Poly-(ADP-ribose) polymerases (PARP) inhibitors are an emerging class of drugs used to treat triple negative breast cancers. The inhibition of PARP, an enzyme involved in DNA repair, used in combination with chemotherapy which induces DNA damage, has been shown to inhibit the proliferative and angiogenic properties of tumours and promote apoptosis [91], as well as improve progression-free and overall survival rates [92].

Chemotherapeutic agents are often used in combination regimens comprising: an immune system suppressant such as cyclophosphamide, platinum compounds such as carboplatin or cisplatin which interfere with DNA replication, taxanes such as paclitaxel and docetaxel that cause cell-cycle arrest and apoptosis, and anthracyclines such as doxorubicin and epirubicin which affect DNA synthesis and repair. A common chemotherapy combination, known as FEC-T, combines fluorouracil (5-FU, an antimetabolite to prevent cell proliferation), epirubicin, cyclophosphamide and docetaxel.

1.4.2 Pathological complete response to neoadjuvant chemotherapy

Pathological complete response (pCR) to neoadjuvant chemotherapy is indicative of a good overall survival [93] and is correlated with long-term outcomes for triple negative and HER2 positive disease [94]. There is still no consensus in the literature on the definition of pCR, though the widely used definition specifies no residual invasive tumour on pathologic assessment in the breast as well as the axillary nodes after surgery [95]. The association of long-term outcomes (such as event-free survival and overall survival) with pCR is challenging due to varying definitions of pCR reported in heterogeneous patient populations [94]. A meta-analysis of 11,000 patients by Houssami et al. showed that though 80% of patients showed some response to neoadjuvant chemotherapy, only 19% of patients overall achieved complete pathological response [96]. Molecular subtype was associated with pCR, with the highest percentages of pCR found in patients with triple negative (31%) and HER2 positive breast cancers (39%) due to the success of targeted therapy. Patients with hormone-receptor-positive (ER+, PR+) breast cancer had the lowest percentage of pCR (8%).

1.4.3 Monitoring response to neoadjuvant chemotherapy

Imaging after neoadjuvant chemotherapy is completed is necessary to assess the extent of residual disease before surgery, aiding the selection of mastectomy, partial mastectomy or lumpectomy and reducing the need for surgical re-excision by improving positive margin rates. Additionally, monitoring response to treatment during the course of chemotherapy also allows for modification of treatment if a patient is not showing therapeutic response or the stopping of treatment if the patient has achieved pCR, avoiding unnecessary toxicity and cost. MRI is the most sensitive technique in the pre-operative assessment of breast cancer [97] and is also the most accurate modality for the assessment of response to neoadjuvant chemotherapy [98–102]. A study by Croshaw et al. found that MRI achieved a positive predictive value (PPV) of 93% and a higher accuracy (84%) than physical breast examination (57%), mammography (74%) or ultrasound (79%) in predicting response to treatment [103].

Tumour size

Reducing the size of a tumour before surgery is a primary aim of neoadjuvant chemotherapy. Changes in tumour size are monitored over the course of therapy, most often noting the size of the longest tumour diameter at each stage. While the gold standard for the assessment of tumour response is the pathological lesion size determined after surgery, a standardised method for measuring response to treatment on imaging was described by the Response Evaluation Criteria for Solid Tumours (RECIST), which defines 4 categories (Table 1.2) [104]. Measurement of the longest tumour diameter is most accurate on MRI compared to mammography and clinical examination [105].

Table 1.2 RECIST criteria for tumour response to treatment [104].

Category	Definition
Complete Response	Disappearance of all target lesions
Partial Response	Reduction in tumour diameter of $\geq 30\%$
Progressive Disease	Increase in tumour diameter of $\geq 20\%$
Stable Disease	Neither sufficient shrinkage to qualify for partial response, or sufficient increase to qualify for progressive disease

However, patterns of tumour shrinkage vary depending on a number of factors, complicating the measurement of tumour size on MRI. A proportion of tumours showing almost complete response do not shrink as a single mass but present as scattered foci of residual

tumour cells in a sea of fibrous tissue, whereas tumours that show little clinical response showed similar morphology to prior to treatment [106].

Therapy induced changes can cause significant over- and under-estimation of residual tumour size. For tumours that exhibit fibrosis and reactive inflammation to treatment, size can be significantly overestimated on MRI [107]. While the high spatial resolution of MRI allows for better visualisation of small lesions and foci than other modalities, the size of tumours that fragment into small foci and scattered residual tumour cells can be underestimated [108]. The use of a chemotherapy regimen containing taxanes, resulting in a diffuse spread of residual disease compared to the single nodular residual lesions seen when using a FEC-based regimen, can lead to an underestimation of residual disease on MRI [108]. The use of antiangiogenic drugs, inhibiting the delivery of MR contrast agents through the tumour microvasculature, can also result in an underestimation of the extent of residual disease [109]. Molecular subtype can impact the determination of residual tumour size and has been shown to be less accurate for luminal type tumours and most accurate for triple-negative and hormone receptor negative/HER2 positive tumours [110, 111], most likely due to the lower contrast uptake of luminal type tumours on MRI [111]. Given the difficulties in accurately determining residual tumour size and margins on imaging, the pathologic assessment of the extent of tumour response from surgical specimens remains the gold standard.

The measurement of tumour volume as opposed to diameter is useful for tumours that have an irregular size and shape, such as invasive lobular carcinomas or tumours with diffuse patterns. From a systematic review of 13 studies, a higher accuracy was found in the prediction of tumour response using tumour volume than a uni- or bi-dimensional measure of tumour size [112]. A study by Hylton et al. found that the strongest predictor of pathologic response was achieved using a combination of MR imaging and functional tumour volume [113]. Functional tumour volume, as measured by comparing the signal enhancement in each tumour voxel from pre-, early and late post-contrast DCE-MRI images, was found to be more accurate than measurements of the longest tumour diameter, as well as a predictor of recurrence free survival [114]. Use of a functional tumour volume is of particular importance for cases which have no residual disease but a remaining necrotic mass [115], or for mucinous carcinoma where a responsive tumour will have a low tumour cellularity but retain a mucin pool visible on imaging [116].

Physiologic Changes

Functional MRI techniques can also be used to monitor physiologic changes which can indicate early tumour response as these changes will occur before a significant reduction in tumour size is observed [115]. Changes in qualitative and quantitative DCE-MRI derived

features indicate changing perfusion characteristics and response to therapy. Woolf et al. showed that changes in signal time-intensity curves were significantly correlated to clinical and pathological response [117]. Kim et al. showed that a ratio of lesion enhancement to background parenchymal enhancement on early-phase DCE-MRI could differentiate pathological complete response from residual disease after chemotherapy [118]. Changes in heuristic [119] and pharmacokinetic [120–123] parameters derived from DCE-MRI have also been found to be significant in the prediction of therapeutic response. The transfer coefficient K^{trans} , a measure of tumour perfusion, has been shown to be a promising early indicator of treatment response [124].

Cellular Changes

Neoadjuvant chemotherapy induces changes at a cellular level, significantly changing the histopathologic appearance of tumours [125]. Studies investigating cellular changes to tumour specimens after chemotherapy found a significant reduction in tumour cellularity [126] and increased nuclear atypia [127], with one study finding it difficult to distinguish between residual tumour cells and chemotherapy induced atypia [106]. Combined with a measurement of residual tumour size, the assessment of response using a measure of tumour cellularity can be more accurate than using tumour size alone [128].

Changes in tumour ADC as measured using DWI reflect these changes in cellularity. Response to treatment has been associated with changes in ADC as measured after one cycle [129, 130], two cycles [129], at mid-treatment [131], and after treatment [131, 132]. Increases in ADC after chemotherapy have also been associated with the presence of necrosis and increasing cell lysis [67].

DWI can be a useful tool in the early assessment of response to therapy as changes in ADC have been shown to occur before reduction in tumour size [133, 134]. Studies by Park et al. and Iaconi et al. found that tumours with lower pre-treatment ADC and high cell density responded better to neoadjuvant chemotherapy [135, 136]. A meta-analysis by Chu et al. found a pooled sensitivity of 88% and specificity of 79% for ADC in the prediction of pCR, though with a higher pooled area under the curve (AUC) when using the change in ADC during treatment than pre-treatment ADC (0.80 vs 0.63) [137].

Radiomics and Machine Learning

The emerging field of radiomics extracts large numbers of quantitative features from medical images. Radiomic analysis is increasingly used in the prediction of pathological complete response in the breast, with studies investigating texture features, shape features, and kinetic features extracted from DCE-MRI, T_2 -weighted images and DWI. These features are used in

combination to create decision support models using machine learning classifiers where the most predictive features are selected. Texture features can be used to quantify the distribution of pixel grey-levels within a tumour volume, reflecting intratumoural heterogeneity, a biomarker of poor prognosis [138]. It has also been shown that the incorporation of radiomics features from the peritumoural region, the background parenchyma, and the contralateral breast can improve the accuracy of prediction [139, 140]. Given that different tumour molecular subtypes respond differently to therapy, studies that carried out separate radiomic analyses based on receptor status found that dividing patients into a hormone receptor-positive group and a triple-negative/HER2+ group improved the accuracy of prediction [139]. The radiomic features most predictive of response have also been shown to vary between different molecular subtypes [139, 141].

While features derived from mid-treatment imaging or changes in features between baseline and mid-treatment imaging have been shown to be the strongest predictors of pathologic response [142, 143], there is growing interest in using baseline imaging to identify predictive imaging biomarkers pre-therapy to improve treatment stratification and provide a more individualised patient prognosis. Radiomics features derived from pre-treatment MRI have been shown achieve a moderate accuracy in predicting treatment response (AUCs 0.69-0.83) [141, 144–147].

Unsupervised machine learning techniques circumvent the need for the extraction of hundreds of radiomics features and subsequent feature selection. Deep learning methods such as convolutional neural networks (CNNs) have not been as widely investigated in the breast for the prediction of treatment response as large datasets are required. Initial studies have used CNNs to predict pCR with mixed results (AUCs 0.55-0.85) [148–151].

1.5 Thesis aims and outline

This thesis focuses on the use of multi-parametric MRI techniques to improve the detection of breast cancer and the assessment of response to neoadjuvant chemotherapy. DWI is a promising non-contrast MRI technique. However, more work is required to improve the image quality and increase the clinical utility of DWI. In this thesis, emerging diffusion models and acquisition strategies in the detection and diagnosis of breast cancer will be assessed. DCE-MRI achieves the highest sensitivity in the detection of breast cancer, though the widespread use of MRI for breast screening is limited by cost and long examination and interpretation times. Abbreviated MRI protocols are able to overcome these limitations. However, a systematic comparison of abbreviated and full diagnostic protocols is required to demonstrate equivalent sensitivity and specificity. Furthermore, while DCE-MRI is

increasingly utilised in the prediction of pathological complete response to neoadjuvant chemotherapy, further investigation is required into the use of radiomics and machine learning approaches to identify predictive imaging biomarkers of response from pre-treatment MRI.

Therefore, the main objectives of this thesis are:

1. To systematically compare models of diffusion, through a meta-analysis, to establish which achieves the highest diagnostic accuracy in the differential diagnosis of breast cancer lesions.
2. To investigate acquisition techniques such as parallel imaging and multi-shot echo-planar imaging that aim to improve upon the image quality of existing clinical diffusion MRI protocols.
3. To systematically compare the performance of abbreviated and full diagnostic protocols in the detection of breast cancer through a meta-analysis.
4. To establish which radiomics features derived from pre-treatment DCE-MRI, using machine learning modelling, are able to predict pathological complete response to neoadjuvant chemotherapy.

Chapter 1 has provided an introduction to breast cancer and the role of imaging. The functional imaging techniques are introduced with reference to the underlying tumour biology they investigate, and neoadjuvant chemotherapy for breast cancer is discussed.

Chapter 2 outlines the core scientific basis on which this PhD is based. The fundamental principles of magnetic resonance imaging are explained, and the pulse sequences used in subsequent chapters are introduced.

Chapter 3 provides a systematic review and meta-analysis of the current models of diffusion. The diagnostic performance of parameters from each model are compared and the variations in methodology of each model in the current literature are assessed.

Chapter 4 describes experiments undertaken to compare a multi-shot diffusion-weighted imaging technique (multiplexed sensitivity encoding, or MUSE-DWI) to conventional single-shot echo-planar imaging (ss-EPI-DWI), investigating the effect of varying parallel acceleration factor and number of shots on quantification of the ADC, distortion and blurring. A refined MUSE protocol is applied to a small cohort of patients and the improvements in image quality and ADC quantification are compared.

Chapter 5 presents a systematic review and comparative meta-analysis of abbreviated and full diagnostic protocol MRI. The diagnostic performances of each protocol are compared in screening and enriched cohorts separately and the variations in patient population, study methodology, and abbreviated protocols are assessed.

Chapter 6 assesses the ability of machine learning models trained using shape, histogram, and texture features extracted from pre-treatment DCE-MRI to predict pathological complete response to chemotherapy. The features selected by logistic regression classification models using the least absolute shrinkage and selection operator are compared, and the predictive performance of features derived from independent time points in the full dynamic series is investigated.

Chapter 7 summarises the main findings of this work and provides outlines for future work.

Chapter 2

Magnetic Resonance Imaging

This chapter describes the fundamental physics of nuclear magnetic resonance and magnetic resonance imaging (MRI). The pulse sequences and acquisition techniques that are used in later chapters are introduced, and the principles of dynamic contrast-enhanced MRI and diffusion-weighted imaging are discussed.

2.1 Nuclear magnetic resonance

2.1.1 Nuclear magnetism

MRI is based on the principle of nuclear magnetic resonance (NMR), which describes the interaction between nuclei and strong external magnetic fields. NMR was first described by Isidor Rabi in 1938 [152] and later developed by both Felix Bloch at Stanford and Edward Purcell at Massachusetts Institute of Technology independently in 1946 [153, 154]. Magnetic resonance has been used for the study of disease since it was discovered by Raymond Damadian in 1971 that the signal from healthy tissue was different from that of cancerous tissue [155]. MRI is now widely used as a non-invasive imaging technique for a wide range of clinical applications.

Atomic nuclei with an odd number of nucleons have a non-zero angular momentum, or ‘spin’. The combination of the spin and positive charge of a nucleus generates a magnetic field. This results in a small magnetic dipole moment, μ , which is related to the spin angular momentum, \mathbf{J} , by

$$\mu = \gamma \mathbf{J} \tag{2.1}$$

where γ is the gyromagnetic ratio, a proportionality constant that is an intrinsic property of a nucleus. The magnitude and direction of \mathbf{J} are determined by the nuclear spin quantum number, I . Values of I and γ for nuclei commonly used in MRI are given in Table 2.1.

Table 2.1 Nuclei used in human MRI studies

Isotope	Abundance (%)	Gyromagnetic ratio γ (MHz/T)	Spin quantum number, I	Relative sensitivity of MRI
^1H	99.99	42.58	1/2	1.000
^3He	0.00001	32.43	1/2	0.442
^{13}C	1.108	10.71	1/2	0.016
^{23}Na	100.0	11.26	3/2	0.083
^{31}P	100.0	17.24	1/2	0.066

In the presence of an external magnetic field, \mathbf{B}_0 , nuclei will experience a torque, $\boldsymbol{\tau}$, given by

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}_0 \quad (2.2)$$

perpendicular to both the field and the magnetic moment, as shown in Figure 2.1. This causes a change in angular momentum perpendicular to the angular momentum, causing the nuclei to precess at an angle θ around \mathbf{B}_0 with a motion similar to a gyroscope. The nuclei precess at a characteristic resonant frequency, known as the Larmor frequency, ω_0 .

In the absence of an externally applied magnetic field, the magnetic moments, or ‘spins’, are randomly oriented. In the presence of a static magnetic field \mathbf{B}_0 , conventionally described as aligned along the z axis, the component of magnetic moment along the direction of the magnetic field is

$$\mu_z = -\gamma I_z = -\gamma m_z \hbar \quad (2.3)$$

where m_z is the magnetic moment quantum number which can take values $-I \leq m_z \leq I$ and \hbar is the reduced Planck’s constant. The number of possible spin states for a nucleus with spin I is equal to $2I + 1$. Hydrogen (^1H), the most commonly used nucleus in MRI, has $I = \frac{1}{2}$ and therefore 2 possible spin states of $\pm\frac{1}{2}$, spin up (parallel with the magnetic field, the lower energy state) and spin down (anti-parallel with the magnetic field, the higher energy state). The energy, E , of each state is given by the vector dot product

$$E = -\boldsymbol{\mu} \cdot \mathbf{B}_0 = -\gamma \mathbf{I} \cdot \mathbf{B}_0 \quad (2.4)$$

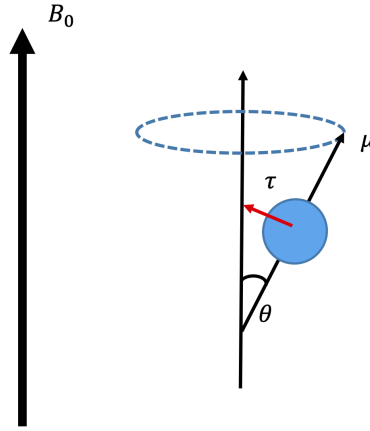


Figure. 2.1 The interaction of a nucleus with a magnetic field. A nucleus with charge and an odd number of nucleons spins about its axis, resulting in a magnetic moment μ . When the nucleus is placed in a static magnetic field, \mathbf{B}_0 , the magnetic moment experiences a torque, τ , and will precess around \mathbf{B}_0 at an angle θ .

The energy difference between the two orientations of spins for $I = \frac{1}{2}$, as shown in Figure 2.2., is given by

$$\Delta E = \left(\frac{1}{2} - -\frac{1}{2} \right) \gamma \hbar B_0 = \gamma \hbar B_0 \quad (2.5)$$

This energy difference has a corresponding frequency from De Broglie's wave equation,

$$\Delta E = \hbar \omega = \gamma \hbar B_0 \quad (2.6)$$

And therefore the resonant frequency, known as the Larmor frequency ω_0 , which corresponds to the only energy at which transitions between the energy levels will be induced, can be written as

$$\omega_0 = \gamma B_0 \quad (2.7)$$

The Larmor frequency can also be given as a scalar frequency, $f_0 = \omega_0/2\pi$. At a typical clinical field strength of 1.5T, the resonant frequency f_0 of a hydrogen nucleus is 64MHz. At 3T, the resonant frequency is 128MHz.

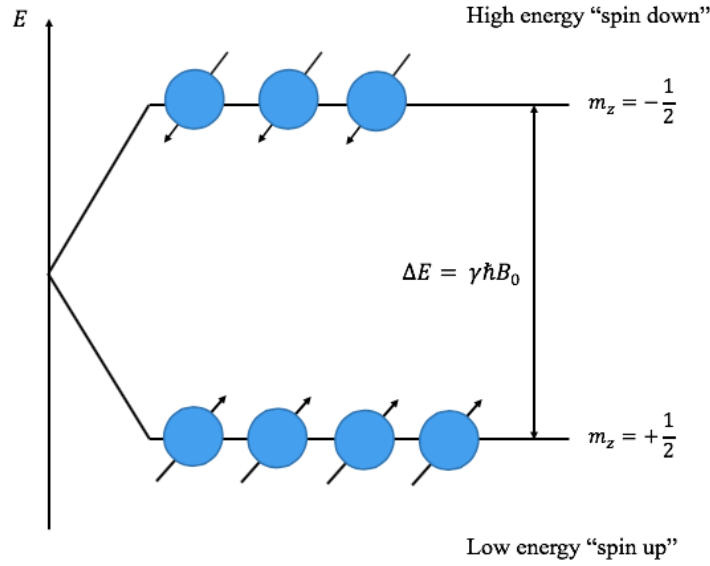


Figure. 2.2 The two possible spin states of nuclei are either oriented spin up (parallel with the field) or spin down (anti-parallel with the field).

2.1.2 Macroscopic magnetisation

The population of spin states across a large ensemble is described by the Boltzmann distribution

$$\frac{N_i}{N} = \frac{e^{-\frac{E_i}{k_b T}}}{Z} \quad (2.8)$$

where N_i is the population of state i , N is the total number of protons, E_i is the energy of the state, k_b is Boltzmann's constant, T is the temperature in Kelvin and Z is the partition function. The ratio of the number of spins in the up state (N_{\uparrow}) and the down state (N_{\downarrow}) can then be written as

$$\frac{N_{\uparrow}}{N_{\downarrow}} = \frac{e^{-\frac{E_{\uparrow}}{k_b T}}}{e^{-\frac{E_{\downarrow}}{k_b T}}} = e^{-\frac{(E_{\uparrow} - E_{\downarrow})}{k_b T}} = e^{-\frac{\gamma \hbar B_0}{k_b T}} \quad (2.9)$$

At clinically relevant temperatures and magnetic field strengths, $\gamma \hbar B_0 \ll k_b T$, and this can be approximated as

$$\frac{N_{\uparrow}}{N_{\downarrow}} \simeq 1 + \frac{\gamma \hbar B_0}{k_b T} \quad (2.10)$$

Therefore, the number of excess spins in the up state is given by

$$N_{\text{excess}} = N_{\uparrow} - N_{\downarrow} = \frac{N_{\text{total}}}{2} \frac{\gamma \hbar B_0}{k_b T} \quad (2.11)$$

This creates a net magnetisation M_0 which is aligned with the B_0 field (Figure 2.3). The net magnetisation per unit volume can be written as

$$M_0 = \mu_z (N_{\uparrow} - N_{\downarrow}) = \frac{\rho_0 \gamma^2 \hbar^2 B_0}{4 k_b T} \quad (2.12)$$

where ρ_0 is the proton density, and given that the magnetic moment of a proton is equal to $\frac{1}{2} \gamma \hbar$. For a magnetic field $B_0 = 1.5\text{T}$ and temperature $T = 37^\circ\text{C}$, there is an excess number of spins in the spin up state, leading to macroscopic magnetisation parallel to the applied field.

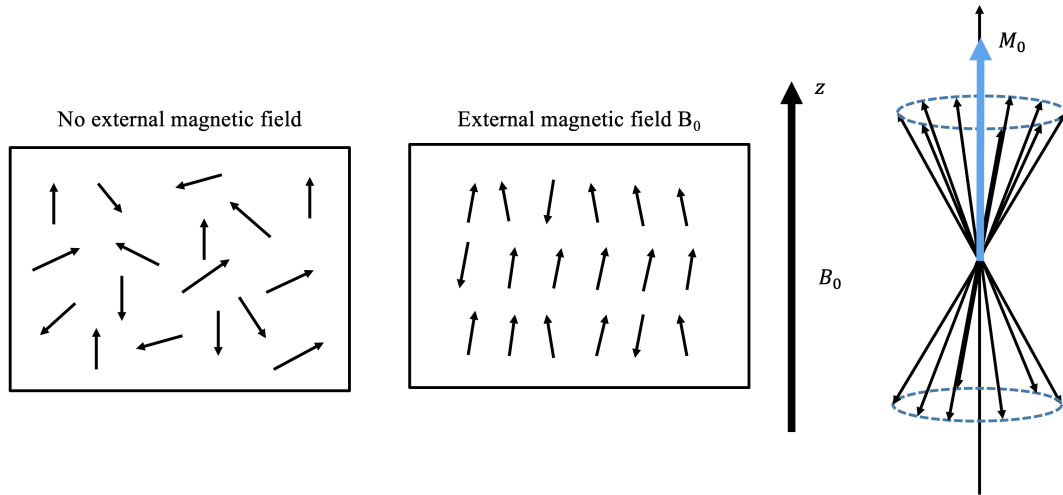


Figure. 2.3 For a large number of nuclei, the excess of spins in the lower energy state gives rise to a net magnetisation M_0 aligned with the magnetic field B_0 .

2.1.3 NMR signal generation

To create an MR signal, the spins must be excited out of equilibrium and the net magnetisation must be displaced out of its equilibrium alignment by an input of external energy into the nuclear spin system. This energy increases the proportion of spins in the higher energy state. A radiofrequency (RF) solenoid coil with its axis perpendicular to the static field, B_0 , creates an oscillating field, B_1 at a frequency ω_1 . When the RF field is switched on at or near the Larmor frequency, known as an RF ‘pulse’, the spins and hence the net magnetisation are rotated into the transverse plane and the net magnetisation will precess. The precession of

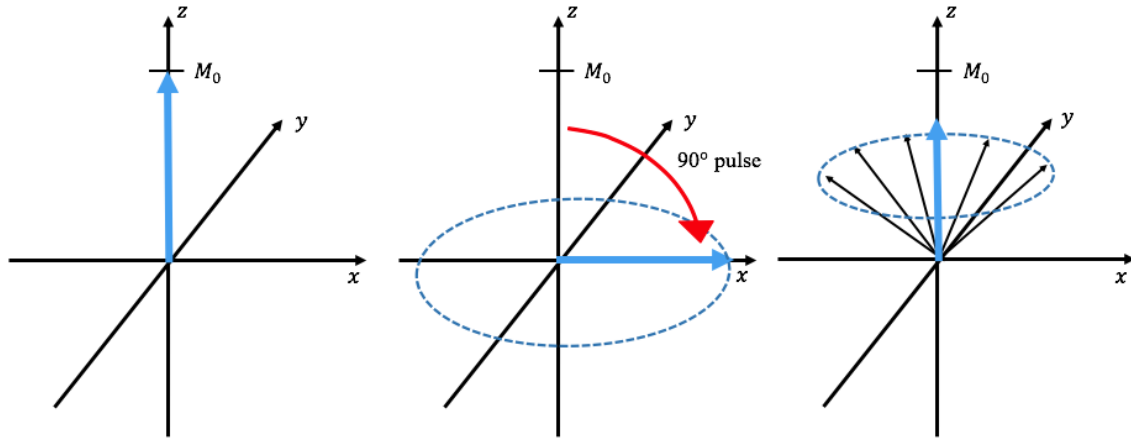


Figure. 2.4 A 90° RF pulse tips the net magnetisation M_0 into the transverse plane. After the RF pulse, the longitudinal component of the net magnetisation recovers to equilibrium and the transverse component of magnetisation decays.

the net magnetisation is known as nuclear magnetic resonance (NMR). The angle through which the magnetisation is rotated, known as the flip angle α , depends on the field strength B_1 and the length of time the field is applied for, t . For a rectangular RF pulse of constant amplitude, the flip angle is given by

$$\alpha = \omega_1 t = \gamma B_1 t \quad (2.13)$$

More generally for pulses of different shapes, the flip angle is given by

$$\alpha = \gamma \int_0^t B_1 dt \quad (2.14)$$

If the magnetisation is rotated perpendicularly into the transverse (x,y) plane, a 90° rotation, this is called a 90° pulse (Figure 2.4). After the RF pulse is switched off, the magnetisation precesses freely and the transverse component of magnetisation decays to zero, returning to equilibrium. This induces an exponentially decaying sinusoidally oscillating voltage in an RF coil (Figure 2.5). This signal is known as the free induction decay (FID).

2.1.4 Mechanisms of relaxation

In his 1946 work, Bloch and his colleagues showed that NMR was a transient phenomenon and posited that the decaying signal can be attributed to the interaction of the spins that make up the net magnetisation with each other and their environment as they return to

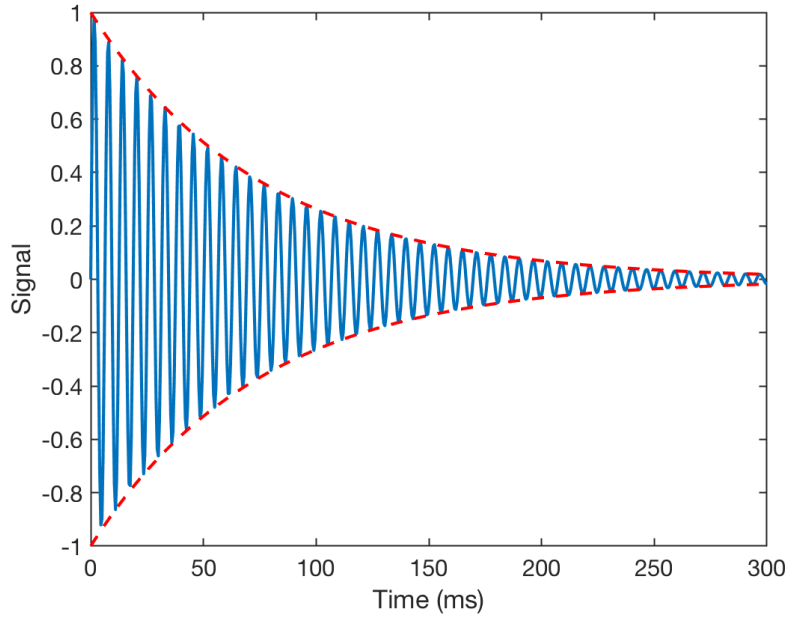


Figure. 2.5 A simulated free induction decay (FID). The signal received in the RF coil oscillates sinusoidally as the net magnetisation precesses at the Larmor frequency and decays by the T_2^* relaxation constant as transverse magnetisation decays to zero and the net magnetisation returns to equilibrium.

thermal equilibrium [153]. The net magnetisation, \mathbf{M} , is a vector rotating in space with three directional components: M_x and M_y , the transverse components (perpendicular to the static field), and M_z , the longitudinal component (parallel to the static field). The Bloch equations are a series of equations using classical mechanics that describe how these three components evolve in time. \mathbf{M} interacts with the magnetic field, \mathbf{B} , by

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B} = \gamma \begin{bmatrix} (M_y B_z - M_z B_y) \mathbf{i} \\ +(M_z B_x - M_x B_z) \mathbf{j} \\ +(M_x B_y - M_y B_x) \mathbf{k} \end{bmatrix} \quad (2.15)$$

where \mathbf{B} includes both the static field along the z axis

$$B_z = B_0 \quad (2.16)$$

and the rotating B_1 field given by

$$B_x = B_1 \cos \omega t \quad (2.17)$$

$$B_y = B_1 \sin \omega t \quad (2.18)$$

The processes by which the longitudinal and transverse components of magnetisation return to thermal equilibrium are known as T_1 and T_2 relaxation, respectively. Therefore, the three directional components of \mathbf{M} can be described by

$$\frac{dM_x}{dt} = \gamma(M_y B_0 + M_z B_1 \sin \omega t) - \frac{M_x}{T_2} \quad (2.19)$$

$$\frac{dM_y}{dt} = \gamma(M_x B_1 \cos \omega t - M_z B_0) - \frac{M_y}{T_2} \quad (2.20)$$

$$\frac{dM_z}{dt} = \gamma(M_x B_1 \sin \omega t + M_y B_1 \cos \omega t) - \frac{M_z - M_0}{T_1} \quad (2.21)$$

For the example in section 2.1.3 where a 90° pulse is applied, these equations have the solutions

$$M_x(t) = M_0 \sin \omega_0 t \cdot \exp\left(\frac{-t}{T_2}\right) \quad (2.22)$$

$$M_y(t) = M_0 \cos \omega_0 t \cdot \exp\left(\frac{-t}{T_2}\right) \quad (2.23)$$

$$M_z(t) = M_0 \left[1 - \exp\left(\frac{-t}{T_1}\right) \right] \quad (2.24)$$

where the M_x and M_y components oscillate at the Larmor frequency ω_0 and decay with time constant T_2 and the M_z component recovers back to M_0 with the time constant T_1 . Calculated T_1 and T_2 relaxation curves for a malignant breast mass are shown in Figure 2.6.

T_1 relaxation, or spin-lattice relaxation, is the process by which the longitudinal component of magnetisation relaxes back to thermal equilibrium. After excitation, energy is exchanged between the excited nuclei and the nearby environment of nuclei and molecules (the lattice) as net magnetisation recovers back to thermal equilibrium, where the spins favour the lower energy (spin up) state.

T_2 relaxation, or spin-spin relaxation, is the process by which the transverse component of magnetisation relaxes back to thermal equilibrium. As precessing spins approach each other and other rotating (tumbling) molecules, the local field experienced by the spins is perturbed. This causes a change in the rate of precession and therefore a loss of phase coherence and

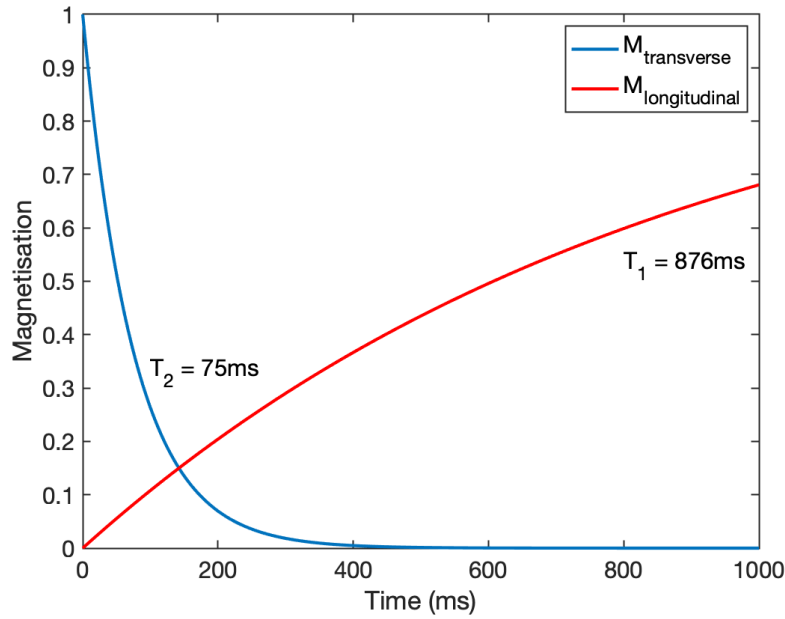


Figure. 2.6 T_1 and T_2 relaxation curves simulated for a malignant breast mass after a 90° pulse [156].

transverse magnetisation decays. In reality, the loss of phase coherence happens more quickly, decaying with time constant $T_2^* < T_2$, such that

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'} = \frac{1}{T_2} + \gamma \Delta B \quad (2.25)$$

where T_2' is the time constant due to field inhomogeneities and ΔB is the field inhomogeneity across a voxel. In practice, this can be compensated for using spin echo pulse sequence, where a refocusing pulse is applied to reverse the effects of inhomogeneities (section 2.3.1).

The rate of relaxation is dependent on the molecular environment of the tissues. Values for T_1 and T_2 for different types of breast tissue are compiled from the literature in Table 2.2.

Table 2.2 In vivo relaxation times in the breast compiled from the literature

	1.5T		3.0T	
	T ₁ relaxation time \pm sd (ms)	T ₂ relaxation time \pm sd (ms)	T ₁ relaxation time \pm sd (ms)	T ₂ relaxation time \pm sd (ms)
Fat	296.01 \pm 12.94 ^a	53.33 \pm 2.11 ^a	366.78 \pm 7.75 ^a	52.96 \pm 1.54 ^a
Fibroglandular tissue	1266.18 \pm 81.8 ^a	57.51 \pm 10.15 ^a	1444.83 \pm 92.7 ^a	54.36 \pm 9.35 ^a
Benign mass	1049.02 \pm 40.31 ^b	89.15 \pm 8.33 ^b		
Malignant mass	876.09 \pm 27.83 ^b	74.76 \pm 3.90 ^b		

^a Rakow-Penner et. al [157]^b Merchant et al. [156]

2.2 Magnetic resonance imaging: hardware and image formation

An MRI scanner has three main components: the static magnetic field, the RF system used to generate and detect MR signals, and the linear gradient coils used to localise the MR signal. A schematic diagram of the MRI system is shown in Figure 2.7.

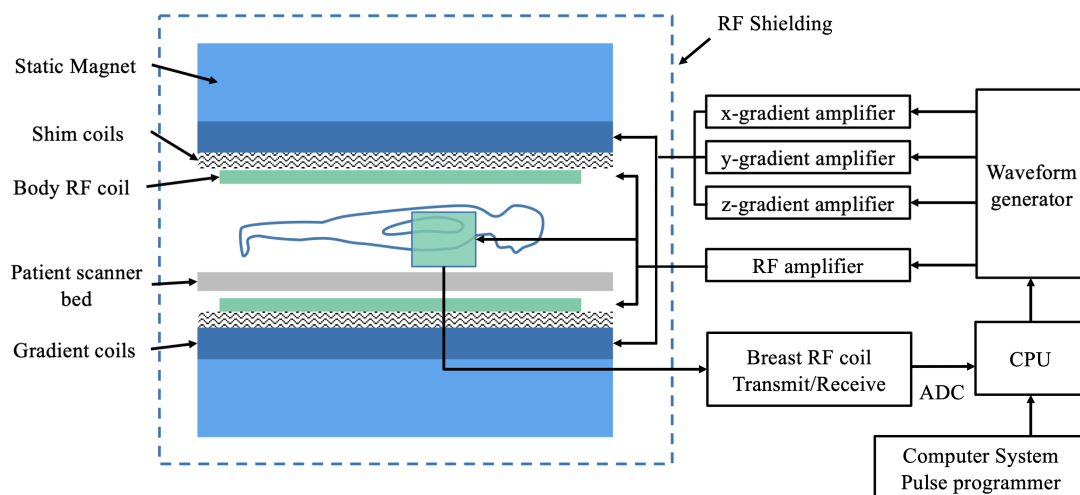


Figure. 2.7 Schematic of an MRI scanner. Patients undergoing breast MRI with a dedicated breast RF coil lie in a prone position.

2.2.1 Static magnetic field

The main homogenous static magnetic field is generated using a superconducting magnet (at $> 1.0\text{T}$). A superconductor has zero electrical resistance at very low temperatures, allowing large electric currents to flow and produce strong magnetic fields. Most commercial clinical scanners operate at 1.5T or 3T . This produces the net magnetisation required for MRI (section 2.1.2).

Most scanners use a magnet comprised of coils of niobium-titanium (NbTi) filaments in a copper matrix. As NbTi is superconductive below 9.4K , the magnet is kept below the critical temperature in a cryostat surrounded by liquid helium (4K) and several insulating and vacuum layers. Inhomogeneities in the field are reduced by shim coils built into the cryostat, and shielding coils protect the environment from unwanted external fields.

2.2.2 Radiofrequency (RF) coils: NMR signal generation and detection

Radiofrequency (RF) coils transmit the B_1 field, or the RF pulse, that is used to excite spins within the subject being imaged. An oscillating B_1 field is produced perpendicular to the static B_0 field by an oscillating electric current passed through the RF-transmit coil. The field is only applied for a short amount of time (as a pulse) and the magnitude of the field can be chosen to change the angle of the rotation of magnetisation into the transverse plane (section 2.1.3). RF transmission is usually carried out using the body coil, a large diameter birdcage coil within the walls of the scanner that is designed to generate a spatially uniform field.

RF coils are also used to detect the MR signal. By Faraday's Law, rotating magnetisation induces a voltage in a receiver coil proportional to the rate of change of magnetic flux. As the magnetisation precesses in the transverse (x, y) plane, a sinusoidally oscillating current is induced in the RF-receive coil. Some coils are built to be able to transmit as well as receive MR signal. In many cases, a smaller coil that is specific to the body part being imaged is used to detect signal, known as a surface coil, with a high sensitivity to RF over a small area which improves the signal-to-noise ratio (SNR).

Specialised coils for breast imaging have been developed that consist of an array of surface coils and are designed to provide bilateral volumetric coverage with a high SNR. Routine clinical breast MRI tends to use 8-channel coils, however 16-channel coils have been developed that offer an improvement in SNR due to the reduction in coil element size compared to an 8-channel coil [158]. Multi-channel coils also allow for the use of parallel imaging, a temporal acceleration technique that combines the spatial sensitivities of the coils to reduce acquisition time while maintaining a high spatial resolution (section 2.3.6).

2.2.3 Gradient coils: spatial localisation

Spatial information in an MR system can be encoded using three orthogonal gradient coils (in the x, y and z directions) controlled independently to produce variations in the magnetic field. These gradients are actively shielded to reduce the effect of eddy currents in the magnet hardware. With only the static magnetic field, all spins in the body precess at the Larmor frequency, ω_0 . To achieve spatial localisation, a linear magnetic field gradient is used to vary the precessional frequency as a function of location along the axis of the gradient.

Slice selection

To achieve slice selection, a linear magnetic field gradient is applied at the same time as the RF pulse in the direction perpendicular to the slice. This causes the resonant frequency to vary linearly with position in the slice select direction e.g. in the z direction

$$\omega(z) = \gamma B(z) = \gamma(B_0 + G_{ss}z) \quad (2.26)$$

where G_{ss} is the magnitude of the slice selective magnetic field gradient. The difference in resonant frequency across the slice is then

$$\Delta\omega = \gamma\Delta B_z = \gamma G_{ss}\Delta z \quad (2.27)$$

where Δz is the slice thickness, as shown in Figure 2.8. An RF pulse centred on the Larmor frequency will excite only the spins in this slice precessing at a frequency within the range $\omega_0 \pm \Delta\omega/2$. This range of frequencies is known as the transmitter bandwidth.

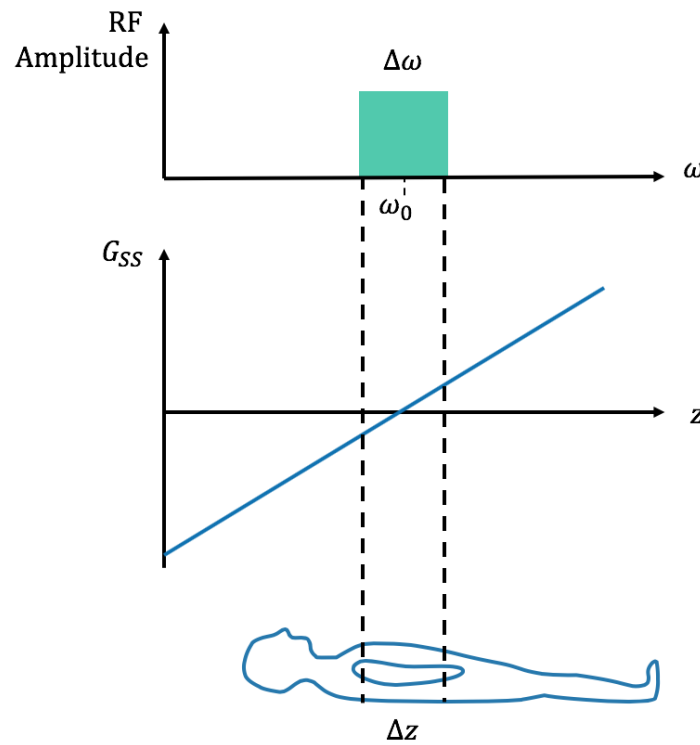


Figure. 2.8 Slice selection. An RF pulse containing a narrow band of frequencies excites a small band of spins. The thickness of the slice Δz can be modified by changing the strength of the slice select gradient as well as the bandwidth of the RF pulse.

To produce a rectangular slice profile in the frequency domain, a truncated sinc pulse is used for the RF pulse. While the slice selective gradient is applied, spins in different positions in the slice will dephase. To rephase the spins, a rephasing gradient of opposite polarity is applied after the slice selective gradient to restore the coherence of the spins.

Frequency encoding

Similarly, if a frequency encoding gradient is applied along the x axis, spins will precess at different frequencies as a function of position along the x axis, creating spatially dependent frequencies (k_x). The frequency encoding ("readout") gradient is applied while the MR signal is acquired. The MR signal measured in the receiver coil is the sum of all the signals with different frequencies from all of the pixels in the slice. Each position along the x direction will correspond to a different frequency, and the amplitude of each frequency is dependent on the density of nuclei in the type of tissue at that spatial location, as shown in Figure 2.9.

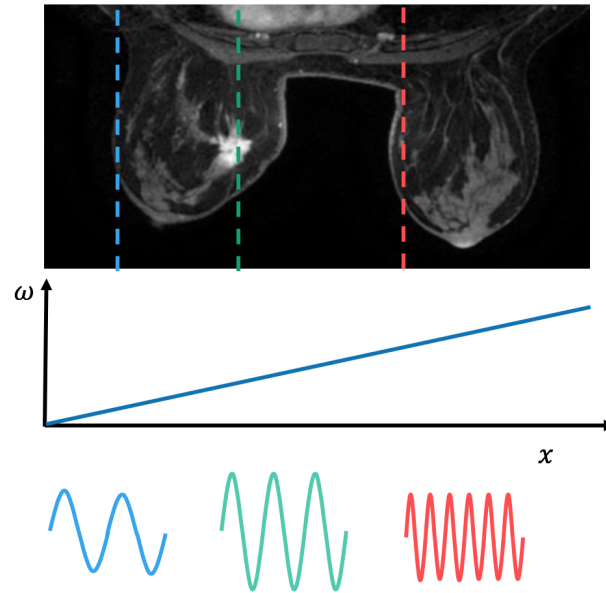


Figure. 2.9 Frequency encoding gradient. A linear magnetic gradient varies the precessional frequency as a function of position, creating spatial frequencies (k_x).

A one-dimensional inverse Fourier transform can be used to separate the individual frequencies from the sum, and the corresponding position of where each signal originated from can be determined in one dimension, as shown in Figure 2.10.

Phase encoding

A phase encoding gradient applied in the y direction is used to encode spatial information in the other dimension by introducing a phase shift between spins that is dependent on location. The phase encoding is carried out before the frequency encoding. When the phase encoding gradient is turned on, the resonant frequencies of the spins temporarily change as a function of position in the y direction. When the gradient is turned off, the spins return to precessing at their original frequencies. However, the spins will have gained or lost phase relative to their state before the gradient was applied, a phase shift ϕ . The magnitude of the phase shift at a location along y is related to the strength of the phase encoding gradient applied, G_{PE} , and the time it is applied for, τ , given by

$$\phi(y) = \int_0^{\tau} \gamma G_{PE}(y) dt = \gamma G_{PE}(y) \tau \quad (2.28)$$

The phase shift will vary as a function of position, creating a pseudo-spatial frequency, k_y , as shown in Figure 2.11. When the frequency encoding gradient is then applied, the MR

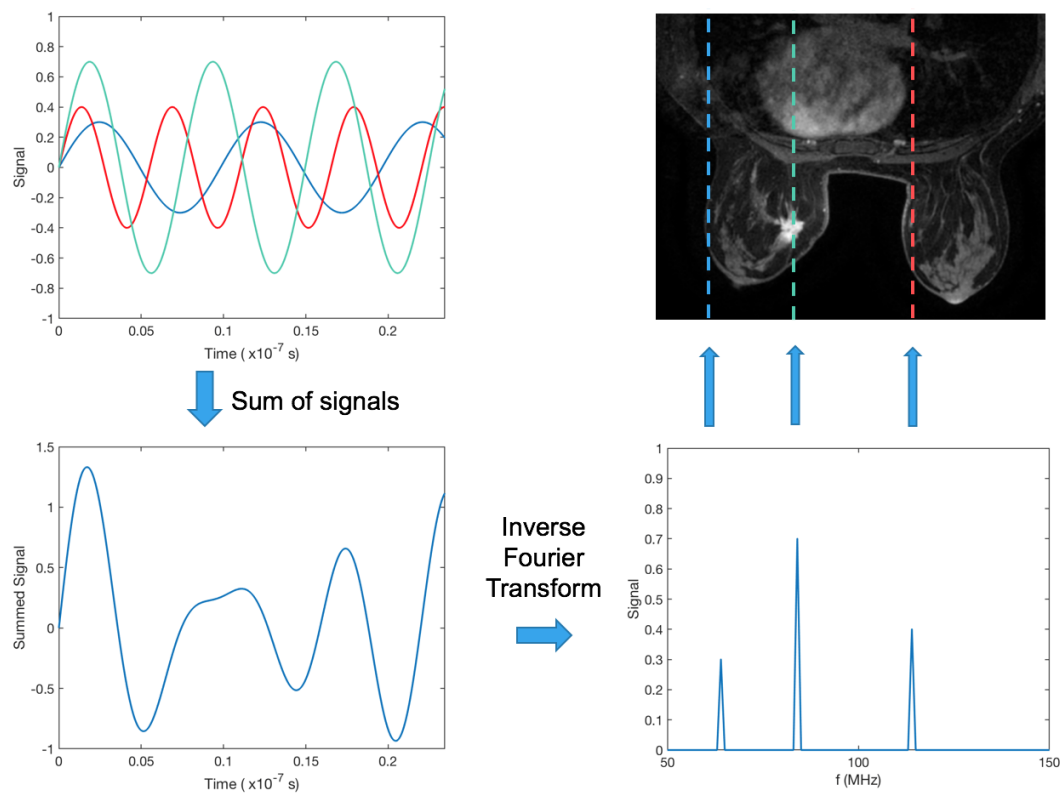


Figure. 2.10 Spatial localisation using frequency encoding. A measured MR signal is the sum of individual signals with frequencies and amplitudes that depend on the location and the type of tissue where they originated from (left). The inverse Fourier transform separates the signals out into spectral lines of specific frequencies which can be mapped back to location using the frequency encoding gradient (right). The amplitude of each spectral line is proportional to the proton density in that location.

signal will include the phase encoding spatial information and every pixel in the image will have a unique combination of frequency and phase. However, it is not possible to separate both frequency and phase information using a Fourier transform from a single MR signal. The phase encoding must be repeated with varying phase encoding gradient strengths, as shown in Figure 2.12, and the different MR signals recorded before a 2D Fourier transform is used to reconstruct the image.

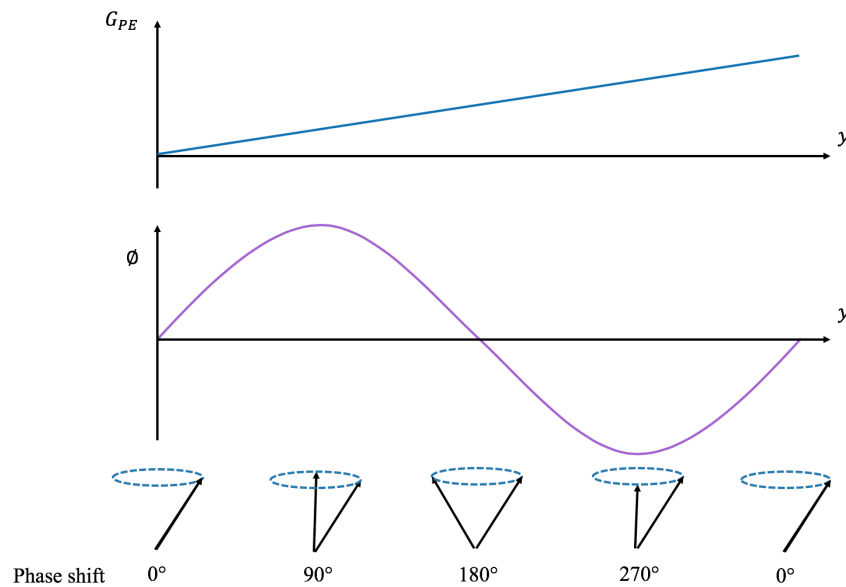


Figure. 2.11 Phase encoding. After the phase encoding gradient is switched on and off, a phase shift, ϕ , is introduced between neighbouring nuclei that varies as a function of position in the y direction.

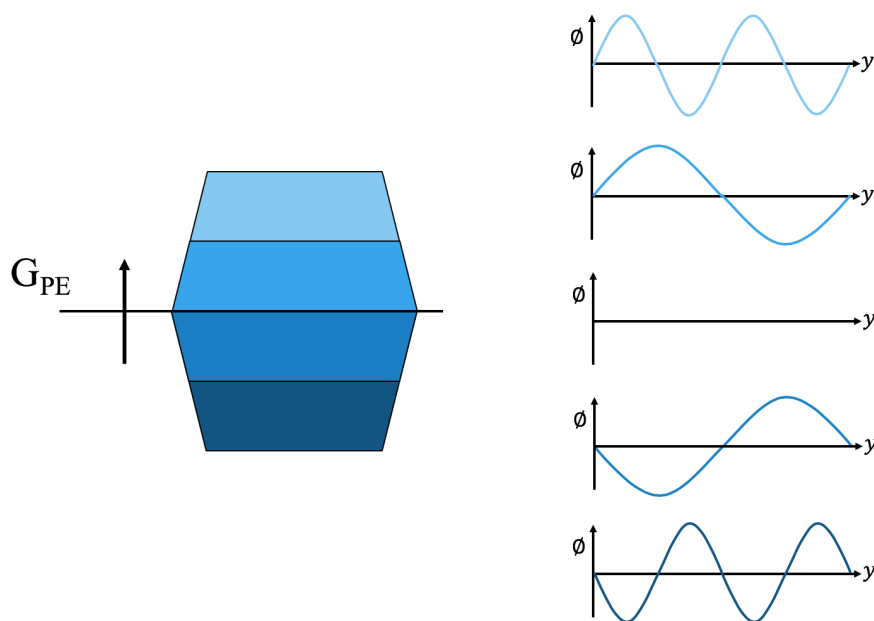


Figure. 2.12 Repeating the phase encoding gradient with different gradient strengths results in a different phase difference across the nuclei.

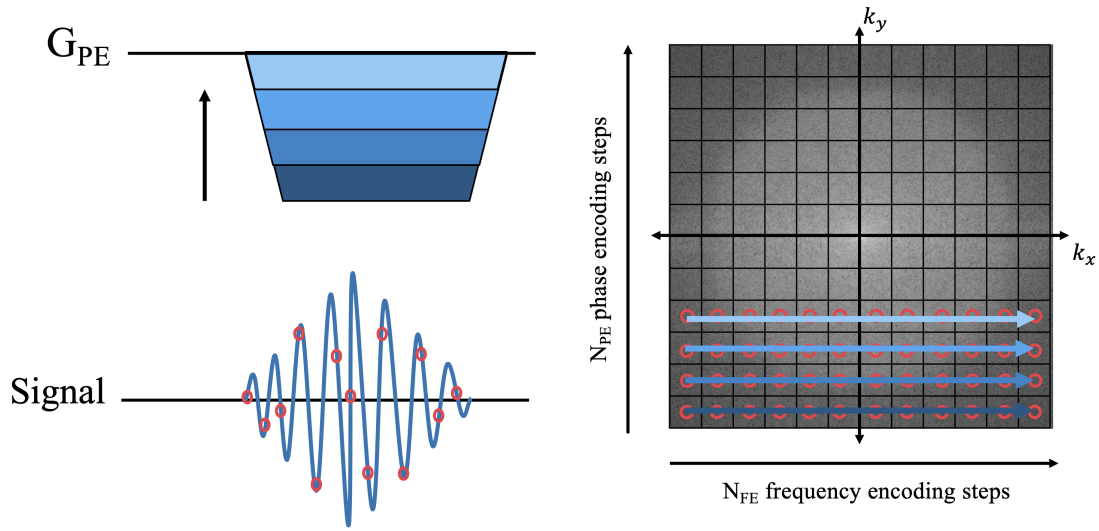


Figure. 2.13 *k*-space diagram. Each time the phase encoding gradient is repeated, the next line of *k*-space is acquired. The number of frequency encoding steps, the number of *k*-space data points in the readout direction, determines the number of samples of the MR signal that are acquired.

2.2.4 Image formation

The MR system is controlled from a workstation, where RF waveforms, gradients and timings can be chosen by a series of control variables. The chosen RF waveforms and gradients are passed to a waveform generator then through the respective gradient and RF amplifiers to generate the required magnetic fields. The signal detected in the receiver coil is amplified, digitised using an analogue to digital converter (ADC), and transferred to a main computer to be processed, as shown in Figure 2.7.

The raw data acquired is stored in a matrix known as *k*-space, representing the spatial frequency components of the image. One phase encoding step is stored as one row in *k*-space (Figure 2.13). The number of data points in *k*-space is given by the matrix size, $N_{FE} \times N_{PE}$, where N_{FE} is the number of frequency encoding steps, relating to the sampling of the MR signal, and N_{PE} is the number of phase encoding steps, the number of times the phase encoding gradient is repeated. When the frequency and phase encoding gradients have been repeated N_{PE} times, all of the information in *k*-space is acquired.

Following the application of the frequency and phase encoding gradients, the elemental signal $\delta S(t)$ at each sampled point is given by

$$\delta S(t) = \rho(x, y) \exp[-i\phi(x, y)] dx dy \quad (2.29)$$

where $\rho(x, y)$ is proton density (taking into account T_1 , T_2 , and T_2^* relaxation). The phase $\phi(x, y)$, according to equation 2.28, is

$$\phi(x, y) = 2\pi \int \gamma G_{FE} x dt + 2\pi \int \gamma G_{PE} y dt \quad (2.30)$$

The total signal over the 2D slice is therefore the integral of equation 2.29 with respect to x and y given by

$$S(t) = \iint \rho(x, y) \exp[-2\pi i (\int \gamma G_{FE} x dt + \int \gamma G_{PE} y dt)] dx dy \quad (2.31)$$

Substituting $k_x(t) = \int \gamma G_{FE} dt$ and $k_y(t) = \int \gamma G_{PE} dt$ gives

$$S(t) = \iint \rho(x, y) \exp[-2\pi i (k_x x + k_y y)] dx dy \quad (2.32)$$

This results shows that the acquired MR signal is the Fourier transform of the proton density $\rho(x, y)$. Therefore a 2D inverse Fourier transform of k -space can be used to reconstruct a 2D image in real space, as shown in Figure 2.14. Lower frequencies, containing the basic contrast information, are stored in the centre of k -space. Higher frequencies, containing higher level detail and edges, are stored at the periphery of k -space. Every point in k -space contains frequency and phase information about every pixel in the image. The in-plane voxel dimensions, Δx and Δy , are governed by the field-of-view (FOV) in each direction and the number of frequency and phase encoding steps.

$$\Delta x, y = \frac{FOV_{x,y}}{N_{FE,PE}} \quad (2.33)$$

2.3 Pulse sequences and acquisition techniques

A single RF pulse that tips the magnetisation into the transverse plane will generate a free induction decay (FID). The FID oscillates at the Larmor frequency and is damped by an exponential decay with time constant T_2^* as the transverse component of the magnetisation dephases. In an MRI pulse sequence, combinations of gradients are played out to manipulate

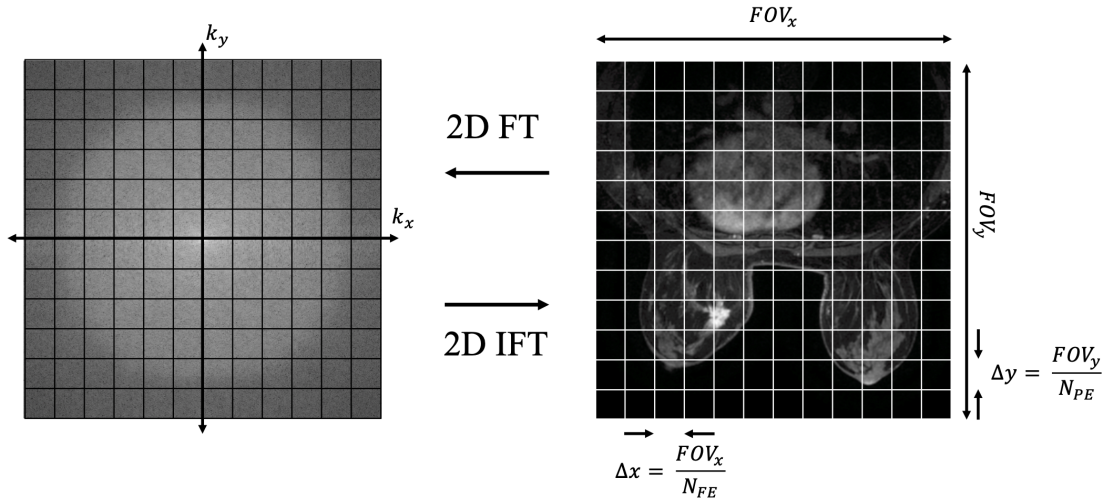


Figure. 2.14 MR image reconstruction. A two-dimensional Fourier transform is used to reconstruct an image from k -space.

the phase of the spins to form a symmetric echo, that is acquired as MR signal. Pulse-timing diagrams are used to graphically represent when gradients and RF pulses are applied.

2.3.1 Spin echo

In a spin echo (SE) sequence, spins are first excited using a 90° RF pulse that tips the magnetisation into the transverse plane. After the 90° pulse, the spins dephase and the transverse component of magnetisation decays due to T_2^* relaxation. A 180° pulse is then applied which flips the magnetisation in the transverse plane, reversing the phase of the spins and refocusing the magnetisation into a spin echo. The refocusing 180° pulse also reverses the effect of the dephasing due to B_0 field inhomogeneities, static tissue susceptibilities, and chemical shifts, such that the echo decays with time constant T_2 as opposed to T_2^* and images will have T_2 contrast. The time between the centre of the 90° pulse and the centre of the echo is known as the echo time (TE). The sequence is then repeated, and the time between the previous 90° pulse and the next 90° pulse is known as the repetition time (TR). Each repeated excitation, the amplitude of the phase encoding gradient is changed to acquire a new line of k -space. A pulse sequence diagram for a SE sequence is shown in Figure 2.15.

Multiple spin echo sequences generate additional echoes without the need for a new 90° pulse. If another 180° pulse is applied before the transverse magnetisation has fully decayed, the spins can be rephased and the magnetisation refocused, forming another echo. In a fast spin echo (FSE) sequence, the 180° pulse is repeated a number of times in one TR period,

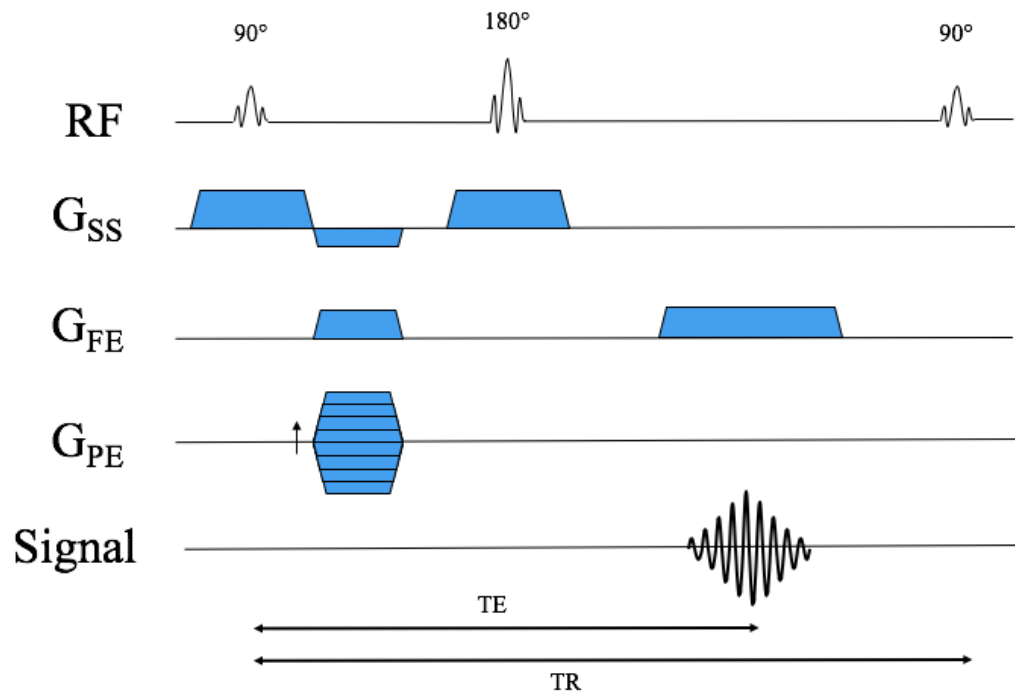


Figure. 2.15 Spin echo pulse sequence. Echo time (TE) is the length of time between the centre of the first exciting RF pulse and the echo and repetition time (TR) is the length of time between RF excitation pulses.

creating multiple spin echoes for which a different phase encoding gradient is applied, such that multiple lines of k -space can be acquired after one 90° pulse, significantly reducing the acquisition time. This is particularly useful for T_2 -weighted images as a long TR period is required to fully recover the longitudinal magnetisation to reduce T_1 -weighting, resulting in long scan times. The number of echoes acquired in one TR period is called the echo train length (ETL). A pulse sequence diagram for a FSE sequence is shown in Figure 2.16.

2.3.2 Gradient echo

In a gradient echo (GRE) sequence, spins are excited using an RF pulse that tips the magnetisation into the transverse plane at an angle given by the chosen flip angle, α (section 2.1.3). A dephasing frequency encoding gradient is applied that accelerates the dephasing of the spins. A rephasing frequency encoding gradient is then applied with the opposite polarity but double the same gradient area that reverses the dephasing, forming an echo as the signal grows then decays. The pulse sequence diagram for a GE sequence is shown in Figure 2.17.

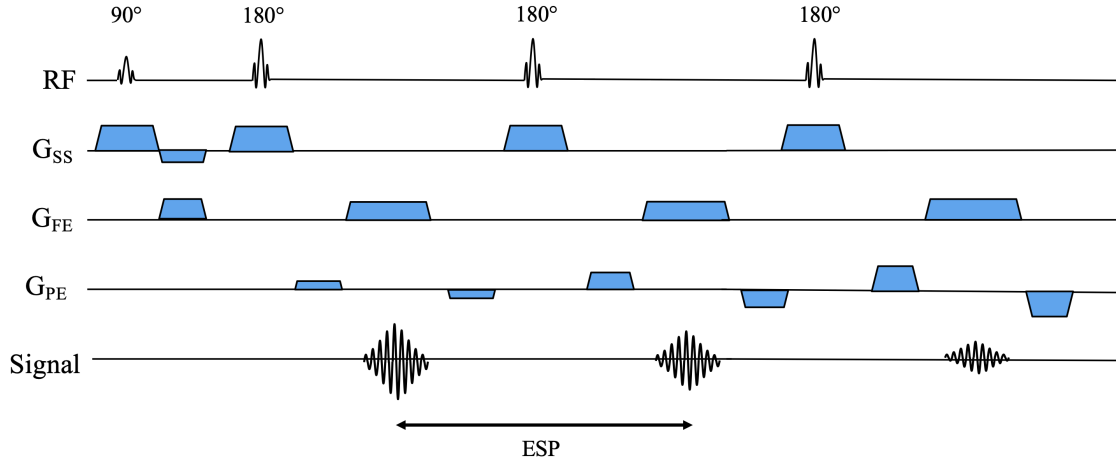


Figure. 2.16 Fast spin echo (FSE) sequence. ESP refers to the echo spacing.

In a gradient echo sequence, only the dephasing of spins due to the applied gradient is reversed and the dephasing due to field inhomogeneities is not reversed. The transverse magnetisation therefore decays due to T_2^* relaxation. For more rapid imaging, flip angles of less than 90° are used so as not to saturate the longitudinal magnetisation, resulting in a shorter TR, and such that the image contrast can be controlled by the flip angle, TE and TR, giving more flexible image contrast. T_1 -weighted gradient echo is used for sequences that require rapid acquisition such as dynamic contrast-enhanced imaging (section 2.4).

Spoiled gradient echo

As TR values used in practice are often shorter than T_2 relaxation times, there is often some residual transverse magnetisation at the end of a TR period before the next RF pulse is applied. Spoiled gradient echo (SPGR) sequences use modifications to the basic gradient echo sequence such that there is zero transverse magnetisation remaining just before the next RF pulse. Strong "crusher" gradients are applied in both the slice select and frequency encoding (readout) directions at the end of a TR period, as shown in purple in Figure 2.17. However, as the gradients produce spatially varying fields, the spoiling is non-uniform. Gradient spoiling is used in conjunction with RF-spoiling, where the phase of the RF-carrier is varied each TR period using a quadratic phase increment [159]. The phase of each subsequent RF-pulse, ϕ_j is given by

$$\phi_j = \frac{1}{2} \phi_0 (j^2 + j + 2) \quad (2.34)$$

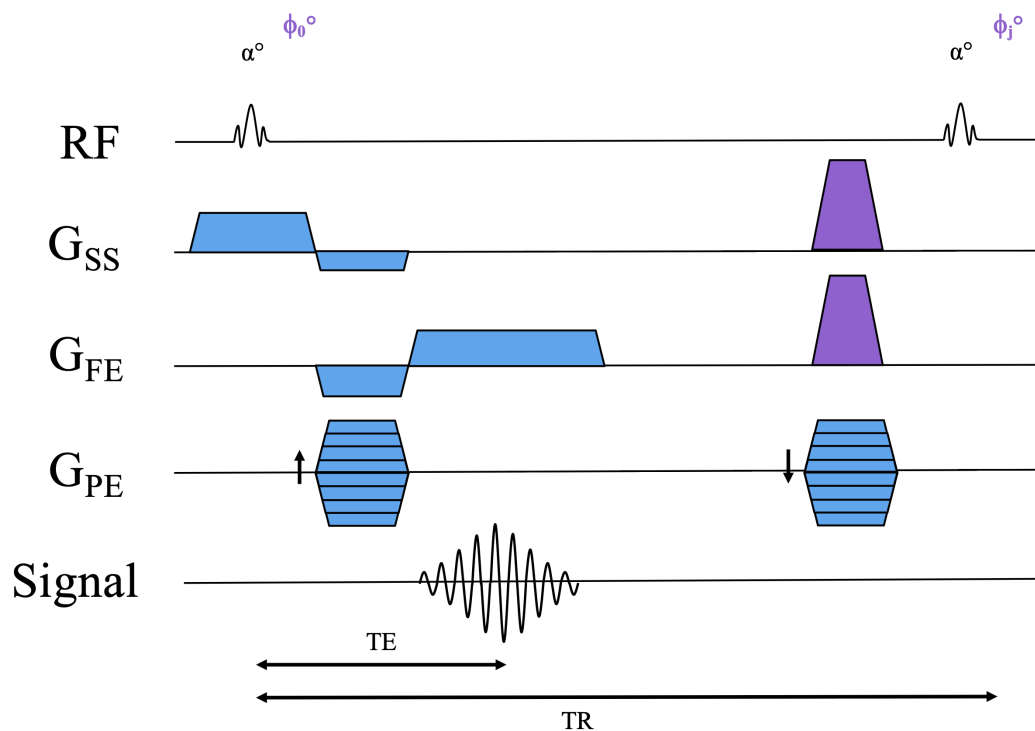


Figure. 2.17 Gradient echo (GRE) pulse sequence. Gradient spoiling crusher gradients and varied RF-carrier phase for a spoiled gradient echo (SPGR) sequence are shown in purple.

where ϕ_0 is the initial RF-spoiling phase. SPGR sequences are generally used to produce T_1 -weighted images and can be used in 2D and 3D.

2.3.3 Generating image contrast

Echo times, repetition times and flip angles can be chosen to manipulate image contrast. Different tissues appear brighter or darker on images that are sensitised to proton density (PD), T_1 or T_2 due to different longitudinal and transverse magnetisation relaxation times. On a T_1 -weighted image, fat is bright as it has a short T_1 relaxation time so longitudinal magnetisation recovers quickly, and it appears hyperintense. On a T_2 -weighted image, fluids are bright as they have a longer T_2 relaxation time, so transverse magnetisation is maintained, and it appears hyperintense. On a proton density-weighted image, tissue with highest concentration of protons (whether it be fat or fluids) will appear hyperintense.

For a spin echo sequence, TR and TE control image contrast. A T_1 -weighted image can be generated using a short TE, as this doesn't give time for T_2 decay differences to appear,

minimising the T_2 contrast, and a short TR to maintain sufficient longitudinal magnetisation to observe T_1 relaxation differences between tissues. Conversely, a T_2 -weighted image can be generated using a long TE to give sufficient time for differences in T_2 decay to appear and a long TR as longitudinal magnetisation will be fully recovered between excitations, minimising the T_1 contrast. A PD-weighted image can be generated with a short TE to minimise T_2 effects and a long TR to minimise T_1 effects.

Similar weighting can be achieved for a gradient echo sequence. However, images will have T_2^* contrast as opposed to T_2 contrast due to the dephasing effects of field inhomogeneities. Flip angles much smaller than 90° can be chosen to maintain longitudinal magnetisation and minimise T_2 effects, reducing the length of the TR period. For a spoiled gradient echo sequence, the signal S can be given by

$$S = k[H] \frac{\sin\alpha(1 - e^{-TR/T_1})}{(1 - (\cos\alpha)e^{-TR/T_1})} e^{-TE/T_2^*} \quad (2.35)$$

where $[H]$ is the proton density and k is a scaling factor. By differentiating with respect to α , the maximum signal can be achieved at the Ernst angle α_E

$$\alpha_E = \arccos(e^{-TR/T_1}) \quad (2.36)$$

While the Ernst angle will give the highest signal for a TR/T_1 combination for a given tissue, it may not necessarily maximise the image contrast between multiple tissues. Suggested flip angles, TR and TE values to generate T_1 -weighted, T_2 or T_2^* -weighted, and PD-weighted images for spin echo and spoiled gradient echo sequences are summarised in Table 2.3.

Table 2.3 Suggested echo times (TE), repetition times (TR) and flip angles for spin echo and spoiled gradient echo sequences to generate different image contrasts, relative to T_1 and T_2 or T_2^* relaxation times.

Contrast Weighting	Spin Echo $\alpha = 90^\circ, 180^\circ$		Spoiled Gradient Echo (TR short $\ll T_1$)	
	TE	TR	TE	Flip Angle α
T_1	Short ($\ll T_2$) <40ms	Short ($\ll T_1$) <750ms	Short ($\ll T_2^*$) <5ms	Large ($>\alpha_E$) 30° - 50°
T_2 or T_2^*	Long ($>T_2$) >75ms	Long ($\approx 5T_1$) >1500ms	Long ($\gg T_2^*$) >30ms	Small ($<\alpha_E$) < 20°
Proton density	Short ($\ll T_2$) <40ms	Long ($\approx 5T_1$) >1500ms	Short ($\ll T_2^*$) <5ms	Small ($<\alpha_E$) < 20°

TE: Echo time, TR: Repetition time, α : flip angle, α_E : Ernst angle

2.3.4 Signal and noise

MR images contain a mixture of signal and noise. Noise, primarily thermal noise, arises from random fluctuations in electrical current, not only in the MR hardware and RF coils used to measure signal but also in human tissue (from electrically charged ions such as sodium, potassium, and chloride, which carry electrical currents within the body). The signal-to-noise ratio (SNR) is defined as the ratio of the signal amplitude to noise level, and is dependent on a number of sequence parameters

$$SNR \propto B_0 \frac{\Delta x \Delta y \Delta z \sqrt{N_{FE} N_{PE} NSA}}{\sqrt{RBW}} \quad (2.37)$$

where NSA is the number of signal averages and RBW is the receiver bandwidth. MR imaging involves choosing imaging options with a trade-off between SNR, contrast-to-noise, resolution and scan time. In order for images to be diagnostically useful, they must have a sufficient SNR to observe small changes in voxel intensity, as well as sufficient contrast-to-noise to observe differences in contrast between different tissues. To improve SNR, k -space can be sampled multiple times (controlled by the NSA) and the resulting data averaged together. To improve spatial resolution, the number of frequency encoding and phase encoding steps can be increased to reduce the voxel size (equation 2.33). However, both of these options result in an increase in scan time. SNR can also be increased by the use of phased-array coils (section 2.2.2), which for N independent coils increases the SNR by a factor \sqrt{N} , although in reality the noise characteristics are not as straightforward.

2.3.5 Fat suppression

MR signal in the body comes from hydrogen nuclei, originating mostly in water and fat. Hydrogen nuclei in water and fat resonate at different frequencies due to their different molecular environments. The difference in resonant frequency of water and fat is 3.5 parts per million (ppm) which corresponds to a frequency difference of 224Hz at 1.5T and 448Hz at 3T. Relative to other tissues, fat has a much shorter T_1 and therefore appears as hyperintense on T_1 -weighted images. A number of techniques have been developed that can be used to suppress or separate water and fat contributions to MR signals, exploiting either their different resonant frequencies, their different T_1 times, or both. Fat suppression is particularly important in the breast due to the large amount of adipose tissue which may obscure pathology or interfere with contrast enhancement.

Short TI inversion recovery (STIR) [160] is a fat suppression technique that exploits the short T_1 of fat. An inversion recovery sequence uses a 180° RF pulse to invert the longitudinal

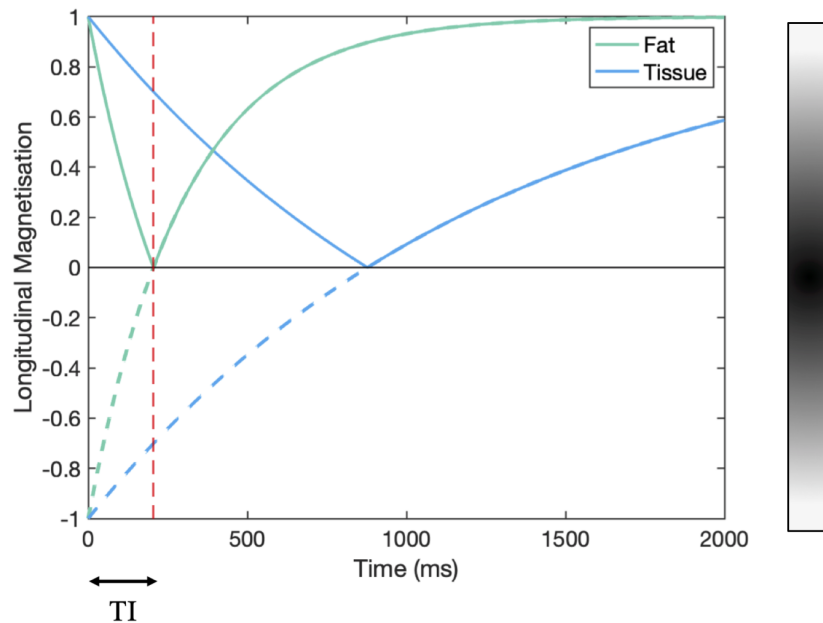


Figure. 2.18 Short TI inversion recovery. An inversion time (TI) is chosen such that fat signal is nulled. For a STIR sequence that uses magnitude reconstruction, the voxel intensity depends only on the magnitude of the signal as opposed to its polarity (solid vs dashed lines). Any tissue with zero magnetisation at time TI appears black in the image.

magnetisation before the rest of the pulse sequence is carried out, where the time between the inverting pulse and the excitation pulse is known as the inversion time (TI). A short TI is chosen such that the signal from the fat is nulled at the time the signal generation for the rest of the sequence takes place (Figure 2.18). Inversion times in the breast are typically 160ms at 1.5T and 250ms at 3T.

Spectrally selective suppression methods, such as Chemical Shift Selective (CHESS) [161], use 90° RF pulses with a narrow bandwidth centred on the resonant frequency of fat applied before the rest of the pulse sequence to saturate the signal from fat before the signal is acquired. Spoiler gradients are then used to dephase the transverse fat magnetisation prior to signal excitation. CHESS requires a homogeneous B_0 field across the FOV such that protons in the fat molecules resonate at the same frequency.

Hybrid techniques such as SPIR (spectral presaturation with inversion recovery) [162] and SPAIR (spectral attenuated inversion recovery) [163] combine both frequency selective and inversion recovery methods. A greater than 90° pulse at the resonant frequency of fat inverts the longitudinal magnetisation of the protons in the fat molecules and an inversion time is chosen such that the fat signal is nulled, allowing for T_1 recovery between the fat saturation RF pulse and the rest of the sequence and therefore a better nulling of fat signal.

Spoiler gradients are also used to destroy the transverse magnetisation of fat. SPAIR uses an adiabatic 180° inversion pulse which is less sensitive to inhomogeneities in the B_1 field, resulting in more homogeneous fat suppression. Both techniques are still sensitive to B_0 inhomogeneities. Hybrid techniques require a longer imaging time than fat saturation or inversion recovery techniques used alone, and have a higher specific absorption rate (SAR, a measure of the amount of power (heat) deposited into tissue) as more RF energy is deposited into tissue than STIR or CHESS.

The Dixon two-point method [164] acquires two images when water and fat nuclei are precessing in phase and out of phase by using two different echo times. Immediately after an excitation pulse, water and fat signals are in phase. However, they begin to dephase as they precess at different frequencies. The difference in frequencies is on the order of 3.4 parts per million (ppm) or 3.4×10^{-6} . This corresponds to a difference in frequency, or chemical shift, of 224Hz at 1.5T and therefore the signals will cycle between being in phase and out of phase with a period $1/224\text{Hz} = 4.5\text{ms}$. At 3T where the chemical shift is doubled, the phase cycling occurs twice as fast with a period of 2.2ms. For an image acquired at an echo time when fat and water are in phase, the signal S_{ip} is the sum of the water signal, S_w , and the fat signal, S_f ,

$$S_{ip} = S_w + S_f \quad (2.38)$$

When the signals are 180° out of phase, the signal S_{oop} is the difference

$$S_{oop} = S_w - S_f \quad (2.39)$$

By adding and subtracting the two images,

$$S_{ip} + S_{oop} = (S_w + S_f) + (S_w - S_f) = 2S_w \quad (2.40)$$

$$S_{ip} - S_{oop} = (S_w + S_f) - (S_w - S_f) = 2S_f \quad (2.41)$$

water-only and fat-only images can be generated.

Spectral-spatial water only fat suppression is a technique that is similar to CHESS but excites only water nuclei [165]. A binomial set of RF pulses are used to select a spatial band (the slice) and a spectral band (at the resonant frequency of water), such that no protons in fat or water molecules outside of the slice are excited. The first RF pulse with half the desired overall flip angle is applied and water and fat are both excited and begin to lose phase coherence. When the water and fat nuclei are precessing 180° out of phase, a second RF pulse is applied such that the net effect of the pulses produces a 90° excitation for water and

a 0° excitation for fat. Water excitation is ideal for rapid imaging sequences as no spoiler gradients or delays to wait for relaxation are required.

2.3.6 Parallel imaging

Parallel acceleration is an image reconstruction technique that reduces acquisition time by exploiting the fact that a receiver coil is made up of multiple receiver channels (section 2.2.2). Lines of k -space are skipped at regular intervals as a function of the acceleration factor R , the ratio between the number of k -space lines sampled and the number of lines in fully sampled k -space. For an acceleration factor of 2, every other line of k -space is acquired. This allows for a reduction the number of phase encoding steps that are acquired per slice and therefore a reduction in overall acquisition time. Reducing the number of phase encoding steps is also beneficial for pulse sequences with a long echo train length, such as echo-planar imaging (section 2.5.1), as this reduces T_2^* blurring.

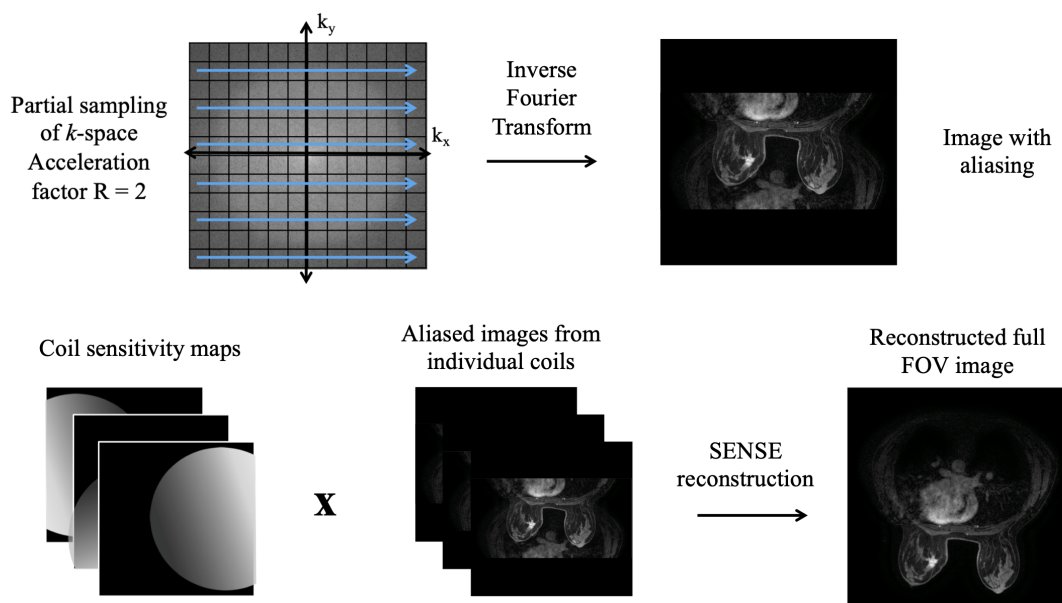


Figure. 2.19 The partial sampling of k -space results in an aliased image, where the anatomy at the edges of the field-of-view is wrapped into the centre of the image. Coil sensitivity maps are acquired from a low-resolution calibration scan then aliased images are reconstructed for each individual coil before SENSE reconstruction is carried out to generate the unaliased full field-of-view image.

Under-sampling of k -space produces reduced FOV images with aliasing as not enough information has been acquired to adequately represent the object being imaged. The spatial variation in the response of each individual coil, the coil sensitivity profiles, are used to assist reconstructing the full un-aliased images by combining data from separate coils. Sensitivity encoding (SENSE) [166] is a parallel imaging method that first reconstructs aliased images from each coil, then uses coil sensitivity maps to separate the superimposed images to reproduce a full FOV image (Figure 2.19). Coil sensitivity maps are generated from low-resolution images acquired using each coil before the main pulse sequence, with intensities normalised using a low-resolution image acquired using the body coil. An image \mathbf{I} with a full FOV can be produced using aliased images \mathbf{P} from coils with sensitivity profiles \mathbf{S} using the equation

$$\mathbf{I} = (\mathbf{S}^H \mathbf{\Psi}^{-1} \mathbf{S})^{-1} \mathbf{S}^H \mathbf{\Psi}^{-1} \mathbf{P} \quad (2.42)$$

where $\mathbf{\Psi}$ is the noise covariance matrix, which describes the correlation of noise in the receiver channels, and the superscript H indicates the transposed complex conjugate [166]. Array coil spatial sensitivity encoding (ASSET) is the GE implementation of SENSE without the acquisition of a body coil calibration image. An alternative parallel imaging technique, known as generalised auto-calibrating partial parallel acquisition (GRAPPA) or auto-calibrating reconstruction for cartesian imaging (ARC), combines data from separate coils in k -space before reconstructing the full FOV image.

2.4 Dynamic contrast-enhanced MRI

Dynamic contrast-enhanced (DCE) imaging is used to acquire a series of T_1 -weighted images before, during, and after the administration of a gadolinium-based contrast agent. Gadolinium based contrast agents shorten the T_1 relaxation time of the blood, making the blood vessels appear brighter on a T_1 -weighted image. The reduction in T_1 is given by

$$\frac{1}{T_{1,post}} = \frac{1}{T_{1,pre}} + R_1 \cdot [C_A] \quad (2.43)$$

where $T_{1,pre}$ is the T_1 of blood before the administration of contrast agent, $T_{1,post}$ is the T_1 of blood following administration, R_1 is the longitudinal relaxivity of the contrast agent and $[C_A]$ is the concentration of the contrast agent in the blood. The contrast agent is administered via a catheter placed within the antecubital vein followed by a saline flush using a power injector for accurate timing and consistency. The series of T_1 -weighted images acquired at regular time points show the passage of the contrast agent through tissue.

While a variety of T_1 -weighted pulse sequences are used to acquire DCE-MRI, a 3D spoiled gradient echo sequence with fat suppression, a short TE and TR, a low flip angle, and parallel acceleration is often used. A trade-off exists between temporal and spatial resolution. High temporal resolution is required to accurately characterise or quantify contrast uptake. High spatial resolution is also required to detect small lesions and characterise lesion morphology. The European Society of Breast Imaging recommends that a typical clinical protocols should use an acquisition time of 60-120 seconds per volume with a minimum in-plane resolution of 1mm^2 [167]. In order to perform pharmacokinetic modelling, a temporal resolution of shorter than 16 seconds per volume has been recommended [168].

Fat suppression is vital for breast imaging due to the high proportion of adipose tissue in the breast. When acquiring DCE-MRI, suppression of fat signal improves the visualisation of the uptake of contrast agent in the fibroglandular tissue, as well as reducing chemical shift artefacts. While imaging at higher field strengths increases the separation of water and fat signal, uniform fat suppression is more difficult due to increased B_1 inhomogeneity. Shimming across both breasts to achieve good B_0 homogeneity is also challenging, limiting the success of conventional fat suppression techniques. Volume image breast assessment (VIBRANT) is a specialised pulse sequence based on 3D gradient echo with modifications to optimise the image quality for breast imaging, using bilateral shims to achieve homogeneous fat suppression across both breasts to enable high quality bilateral imaging.

Three-dimensional imaging

DCE-MRI is often performed in three-dimensional (3D) mode, using an additional phase encoding gradient in the slice select direction, as shown in red in Figure 2.20.

As the sequence is repeated, both the phase encoding gradients are changed and k -space lines in the k_y and k_z direction are acquired, producing a k -space matrix with three dimensions. The region at the centre of k -space is often acquired first, known as centric filling. The 3D image can be reconstructed by applying a 3D inverse Fourier transform to the k -space matrix.

2.4.1 Temporal acceleration

The time-resolved imaging of contrast kinetics (TRICKS) technique [169, 170] can be used to achieve a high temporal resolution by segmenting k -space into a central region and surrounding concentric regions and interleaving the acquisition of k -space points from central and peripheral regions, reducing the number of k -space data points acquired at each time point and therefore reducing the overall acquisition time. Given that the majority of contrast lies in the centre of k -space, the central region is sampled most frequently, and the peripheral

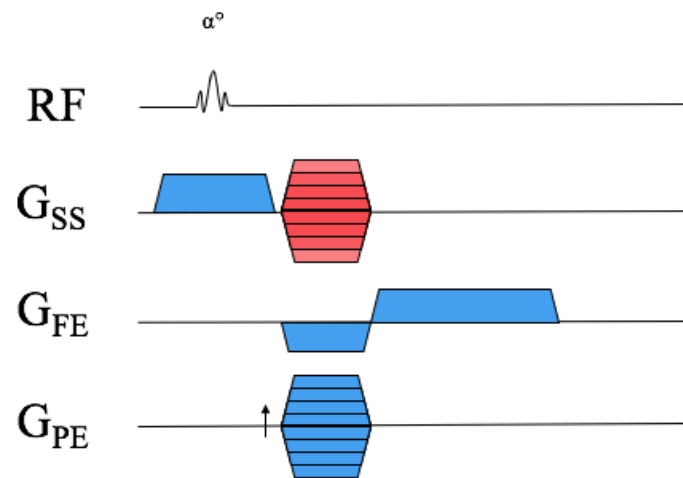


Figure. 2.20 A gradient echo pulse sequence adapted for 3D imaging.

regions are sampled periodically, as shown in Figure 2.21. Through a process known as view-sharing, the k -space data from different time points is combined to reconstruct an image of high temporal and sufficient spatial resolution. Using TRICKS, a temporal resolution of 10s can be achieved to facilitate pharmacokinetic modelling.

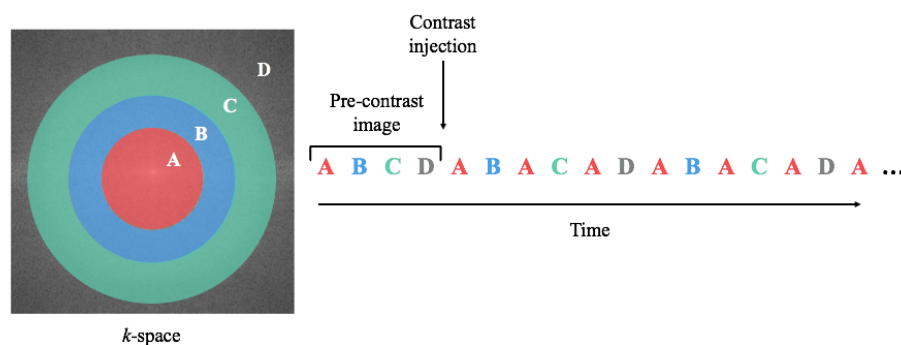


Figure. 2.21 TRICKS acquisition. Before the contrast injection, at least one full set of k -space data, or one pre-contrast image, is acquired. Directly after the administration of contrast agent, the centre of k -space is sampled to sufficiently capture the period of maximum change in contrast enhancement. The other regions of k -space are periodically sampled.

2.5 Diffusion-weighted imaging

2.5.1 Diffusion-weighted preparation

Diffusion-weighted images (DWIs) are acquired using a pulsed gradient spin echo (PGSE) sequence (Figure 2.22), developed by Stejskal and Tanner [171], in which a pair of diffusion sensitising gradient pulses are added to a spin echo sequence on either side of the 180° RF pulse. The amount of diffusion weighting is determined by the amplitude and timing of the diffusion gradient pulses, described by the b-value (units s/mm²)

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right) \quad (2.44)$$

where γ is the gyromagnetic ratio, G is the gradient amplitude, δ is the gradient pulse duration and Δ is the time between the centres of the gradient pulses. In reality, the b-value will also be affected by slew rate, given by

$$\text{Slew rate} = \frac{\text{Peak gradient strength}}{\text{Rise time}} \quad (2.45)$$

where the rise time is the time taken for the gradient to achieve the maximum gradient amplitude required (Figure 2.22). The first of these gradient pulses introduces a phase shift to the water protons dependent on location. The second of the gradient pulses (with the same magnitude as the first gradient pulse but with an opposing phase shift effect) will completely re-phase the protons if they remain in the same location. As the water protons move randomly via Brownian motion and change direction in the time between the diffusion gradients, the protons will experience a different gradient field and the accumulated phase changes lead to a loss of signal within each voxel. The signal attenuation is described by

$$S(b) = S(0) \exp(-b \cdot \text{ADC}) \quad (2.46)$$

where $S(b)$ is the signal with diffusion weighting, $S(0)$ is the signal without diffusion weighting, and ADC is the apparent diffusion coefficient. A series of DWIs can be acquired with two or more b-values by changing the diffusion gradient amplitude and ADC maps calculated by fitting signal decay for each voxel. In tissue, as well as the Brownian motion of water protons, there are a number of other biological processes such as blood and lymphatic flow in the microvasculature and diffusion restriction from microstructures that contribute to signal attenuation. Bulk flow and motion can also affect measurement of signal, and therefore the measured diffusion coefficient is referred to as the 'apparent' diffusion coefficient.

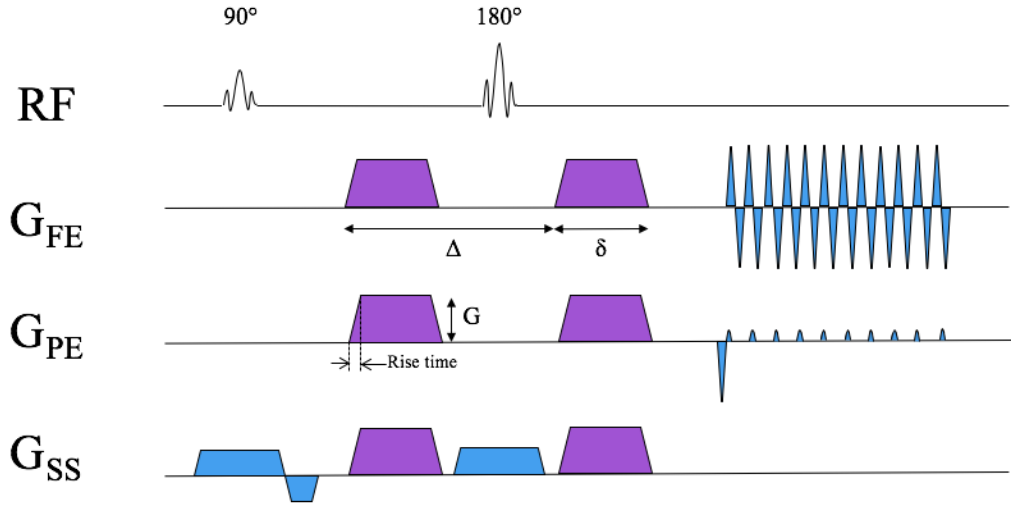


Figure. 2.22 Diffusion-weighted imaging pulse sequence. The spin echo echo-planar imaging sequence is modified with a pair of diffusion-weighting gradients either side of the 180° pulse, shown in purple. Diffusion gradients are applied in 3 orthogonal directions, either individually, simultaneously or in linear combinations of the 3 directions. The gradient amplitude, G , the gradient pulse duration, δ , and the time between the centres of the gradient pulses, Δ , are used to control the diffusion weighting.

Diffusion in biological tissues is anisotropic due to restrictions from membranes and microstructures, and as such the diffusion properties of a tissue can be described as a function of direction by a diffusion tensor

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \quad (2.47)$$

where the three orthogonal elements D_{xx} , D_{yy} , and D_{zz} describe the components of diffusion along each of the three principal directions (x , y , z), and the other six terms describe the correlation in diffusion between these directions. In a standard DWI acquisition, the PGSE sequence is repeated with the diffusion gradients applied in the three orthogonal directions to obtain a directionally invariant measure of diffusion. From the diffusion tensor, the apparent diffusion coefficient can be found from the three orthogonal elements

$$ADC = \frac{1}{3} (D_{xx} + D_{yy} + D_{zz}) \quad (2.48)$$

2.5.2 DWI artefacts

DWIs are normally acquired using single-shot echo-planar imaging (ss-EPI), a modification of a spin echo sequence where the frequency encoding gradient switches rapidly between positive and negative amplitude as the phase encoding gradient is repeatedly blipped, as shown in Figure 2.22. This corresponds to the encoding of one line in k -space, or one phase encoding step, each time the frequency encoding gradient changes. The number of k -space lines acquired in one excitation is known as the echo train length (ETL). When all of k -space is acquired from one single RF excitation this is referred to as a single 'shot'. EPI is often used for diffusion-weighted imaging (DWI) to maintain a high signal-to-noise ratio (SNR) efficiency while minimising the effects of patient motion. DWI is prone to many artefacts due to the EPI readout, particularly at higher field strengths. The main artefacts in DWI are blurring caused by T_2^* decay during the readout, geometric distortion, N/2 ghosting caused by the switching of the bipolar frequency encoding gradient, and chemical shift artefacts.

For an EPI readout with a long echo train, signal decays due to T_2^* relaxation, reducing the signal at the edges of k -space that correspond to fine spatial detail, resulting in blurring. This can be quantified using the point spread function (PSF), which describes the blurring of signal into surrounding voxels. Geometric distortions occur in areas with local B_0 field inhomogeneities caused by poor shimming and variations in magnetic susceptibility, particularly at the interface between air and tissue. These field inhomogeneities affect the accuracy of the small phase encoding gradients, resulting in phase errors and signal being mapped to the wrong location. The size of the susceptibility artefact is given by

$$\text{Susceptibility artefact size} \propto \frac{\Delta \text{Susceptibility} \cdot B_0 \cdot TE}{BW_{PE}} \quad (2.49)$$

where BW_{PE} is the phase encoding bandwidth [172]. The rapid switching of the frequency encoding gradients induces eddy currents in conductive metallic structures of the magnet which also results in geometric distortions in the image. While gradient pre-emphasis and specialised hardware (such as secondary coils used to dynamically cancel magnetic flux changes in the main magnet structure) are often used to mitigate the effect of eddy currents, these effects are more severe for the strong diffusion-weighting gradients used, particularly at large b -values where the gradient amplitude is high. Distortion will vary between images of different b -value, resulting in blurred ADC maps and inaccurate quantification of the ADC.

The back and forth trajectory through k -space can also result in the accumulation of phase shift from eddy currents, off resonance frequencies and other sources such as inadequate shimming and gradient nonlinearity that oppositely affect odd and even echoes as the directionality of the readout is changed. This effect acts as a modulation with a frequency of

1/2 the phase encode bandwidth, and in the Fourier transformed image appears as a replicate reduced intensity image in the phase encode direction shifted by half the FOV, known as "N/2 ghosting".

Chemical shift artefacts arise from fat resonating at a slightly lower frequency than water (section 2.3.5), manifesting as a shift in spatial location in the Fourier transformed image. This is especially apparent for an EPI readout due to the low bandwidth in the phase encode direction (around 1kHz). The shift in voxels in the phase encoding direction, Δy ,

$$\Delta y = \frac{\Delta f_{CS}}{BW_{PE}} \cdot FOV_{PE} \quad (2.50)$$

where Δf_{CS} is the chemical shift. At 1.5T where the chemical shift is about 220Hz, the shift in voxels for a FOV of 360mm is 67.2mm. At 3T where the chemical shift is twice as large, the shift in voxels is twice as large. This is of particular importance for breast imaging due to the large proportion of fat but can be avoided with good fat suppression.

DWI artefacts are mostly mitigated by shortening the overall readout, which reduces the length of the echo train and therefore the amount of T_2^* decay and phase accumulation that can occur. A number of techniques are used that increase the speed of k -space traversal along the phase encoding direction, increasing the phase encode bandwidth, which reduces the size of both susceptibility-related distortions (equation 2.49) and chemical shift artefacts (equation 2.50). Most commonly, parallel imaging is used to accelerate acquisition and under-sample k -space, at the cost of a lower SNR (section 2.3.6).

2.5.3 Multi-shot imaging

Multi-shot techniques can also be used to acquire k -space in a number of segments each with a reduced echo train length and shortened echo spacing. However, there is an increase in overall scan time and increased sensitivity to physiological motion between shots. Examples of k -space trajectories for DWI acquisition are shown in Figure 2.23.

DWI is extremely sensitive to patient motion as diffusion-weighting gradients are sensitive to the diffusion of water on a molecular scale. As such, any bulk patient motion such as respiration or cardiac pulsation will result in phase errors. Readout-segmented EPI techniques, such as RESOLVE (readout segmentation of long variable echo trains) [173], segment k -space in the readout (k_x) direction and acquire navigators (unaliased and low-resolution images) to correct for motion-induced phase errors between shots. The use of navigators reduces the overall acquisition efficiency as each navigator is acquired in 30-40ms and increases the specific absorption rate (SAR) of the sequence by 30% [174].

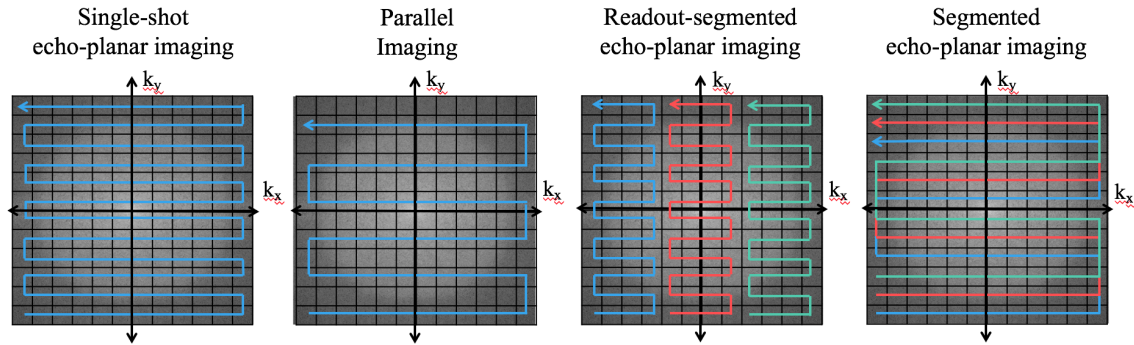


Figure. 2.23 k -space trajectories used for diffusion-weighted imaging: single-shot echo-planar imaging (ss-EPI), ss-EPI using parallel imaging with an acceleration factor of 2, and readout-segmented EPI and segmented EPI with 3 segments.

Phase segmented EPI techniques, such as MUSE [175], acquire interleaved segments in the k_y direction and correct for motion-induced phase errors between shots without the need for navigators. The MUSE technique is an extension of the parallel imaging technique SENSE (section 2.3.6), where the sensitivities of the different channels of the receiver coil are used to help reconstruct the image. After the segments are acquired, navigators are generated using parallel imaging, combining the k -space data from each receiver coil channel to produce a full k -space navigator image. The segments are then corrected using the navigator and combined to produce the full FOV image. While the navigator free method has advantages in terms of scan efficiency and SAR, it is limited by coil geometry as the number of segments acquired is constrained by the number of receiver coil channels.

2.5.4 Advanced diffusion models

Advanced diffusion models attempt to describe the processes of diffusion and probe aspects of the tumour microenvironment not captured by the ADC model, such as perfusion in the microvasculature and diffusion anisotropy (section 1.2.3). These models are able to provide more information about the physiological characteristics of a tumour to enable a better understanding of the mechanisms of cancer and tumour growth.

Non-Gaussian diffusion

Signal decay with b -value has been shown to deviate from the mono-exponential ADC model of diffusion [176], as shown in Figure 2.24. Various non-Gaussian diffusion models have been proposed to achieve an improved fitting of signal decay by using additional parameters to better describe diffusion in biological tissue.

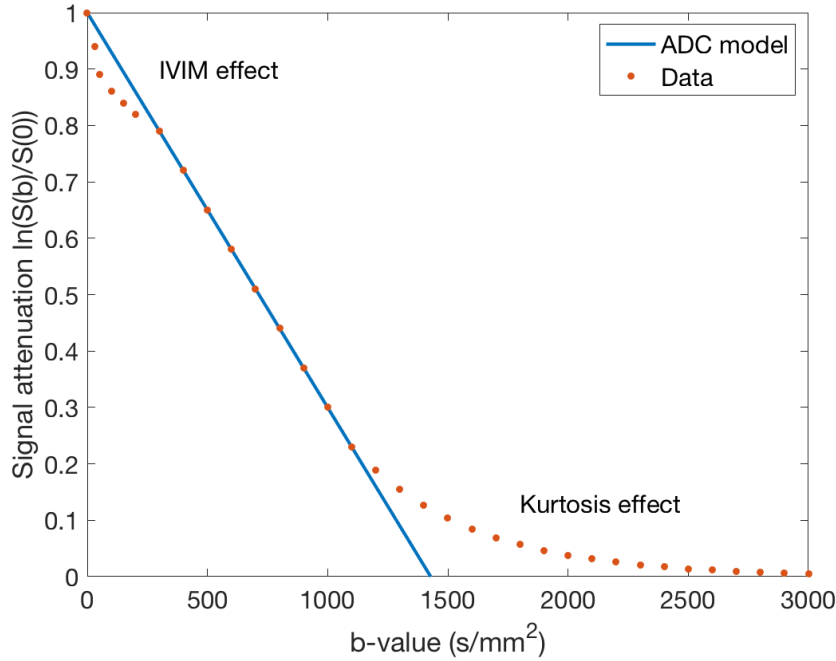


Figure. 2.24 Non-Gaussian diffusion. Signal attenuation with b-value deviates from the ADC model at low b-values ($< 200 \text{ s/mm}^2$) and high b-values ($> 1500 \text{ s/mm}^2$).

Intra-voxel incoherent motion (IVIM), first proposed by Le Bihan et al. [49], expands upon the mono-exponential ADC model and fits a bi-exponential model to the signal decay. While signal decay in DWI can mainly be attributed to diffusion, at lower b-values with weaker diffusion gradients, microcirculation of blood in the capillaries contributes more significantly to signal decay. Signal attenuation is described by

$$S(b) = S(0)[f \cdot \exp^{-D^*} + (1 - f) \cdot \exp^{-D}] \quad (2.51)$$

where f is the perfusion fraction with values between 0 and 1, corresponding to the percentage of voxel volume occupied by capillaries. The tissue diffusivity, D , and the pseudo-diffusivity from the microvasculature, D^* , separate tissue into two compartments with different diffusion coefficients. IVIM modelling requires a wide range of b-values to be acquired, particularly at low b-values where the IVIM mechanism is more prominent, and uses a non-linear least-squares fitting to fit parameters to the bi-exponential model.

Diffusion kurtosis imaging (DKI) probes the non-Gaussian behaviour of diffusion. At high b-values ($2000 - 3000 \text{ s/mm}^2$), signal attenuation has been shown to behave non-linearly [53] and can be described by a quadratic order expansion to the ADC model given by

$$S(b) = S(0) \cdot \exp \left(-b \cdot ADC + \frac{1}{6} b^2 \cdot ADC^2 \cdot K \right) \quad (2.52)$$

where K is the kurtosis parameter.

The stretched exponential model is also used to determine the non-Gaussianity of diffusion, modelling deviation from the mono-exponential model as

$$S(b) = S(0) \cdot \exp(-b \cdot DDC)^\alpha \quad (2.53)$$

where DDC is the distributed diffusion coefficient and α is a stretching parameter, quantifying the degree of non-Gaussianity.

Diffusion tensor imaging

Diffusion tensor imaging (DTI) measures the anisotropy of diffusion due to restrictions from membranes and microstructures and requires at least 6 diffusion gradient directions to probe the extent of diffusion in each direction which are expressed as the diffusion tensor (equation 2.47). The measured values for each component of the diffusion tensor are dependent on the frame of reference that they are measured in, the choice of x , y , and z directions. A spatially invariant alternative co-ordinate system uses the axes of a diffusion ellipsoid determined by the eigenvectors (representing the directions of diffusion) and eigenvalues (representing the magnitude of diffusion) of the diffusion tensor. The main axis of the ellipsoid is parallel to the eigenvector with the eigenvalue of the largest magnitude, the prime diffusion coefficient λ_1 . The other eigenvalues are labelled in order of descending magnitude ($\lambda_1 > \lambda_2 > \lambda_3$). The eigenvalues can be used to calculate parameters such as the mean diffusivity, MD, given by

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (2.54)$$

and anisotropy indices that represent the differences between the eigenvalues, such as the maximal anisotropy index, $(\lambda_1 - \lambda_3)$, and the fractional anisotropy, FA, given by

$$FA = \sqrt{\frac{1}{2} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}} \quad (2.55)$$

Chapter 3

A meta-analysis of the diagnostic performance of diffusion MRI for breast lesion characterisation

Diffusion-weighted imaging (DWI) is increasingly used in the detection and diagnosis of breast cancer. This chapter systematically reviews the literature of all relevant DWI techniques used in the breast and a meta-analysis to assess the diagnostic performance of each technique in the characterisation of breast lesions is performed. Diffusion-weighted imaging, intra-voxel incoherent motion (IVIM) and diffusion tensor imaging (DTI) are shown to have a comparable diagnostic accuracy with a high sensitivity and specificity. However, to date the number of studies investigating advanced DWI techniques are limited and there is a lack of standardisation in methodology.

Contents of this chapter have been published in *Radiology* [50] and presented at the International Society for Magnetic Resonance in Medicine (ISMRM) 2019 (abstract #1853).

3.1 Introduction

MRI has very high sensitivity but lower specificity (93% and 71%, respectively) [10] for the characterization of breast lesions. In order to improve specificity, the diffusion properties of breast lesions have been investigated. The apparent diffusion coefficient (ADC), measured using diffusion-weighted imaging (DWI), is increasingly used as a marker in the detection and characterization of breast lesions. The ADC of a malignant lesion is lower than that of a benign lesion due to the restricted diffusion in regions of high cellular density- often the result of proliferation of glandular tissue. Multiple studies have set a threshold value for ADC

and assessed the diagnostic utility in distinguishing malignant and benign lesions. Advanced diffusion models attempt to capture more complex aspects of the tumor microenvironment, such as diffusion tensor imaging (DTI) which captures diffusion anisotropy, and intra-voxel incoherent motion (IVIM) which captures deviation from the mono-exponential model due to perfusion effects (section 1.2.3).

While previous meta-analyses have assessed the performance of the ADC model in differentiating between benign and malignant lesions [38, 177, 178], more advanced diffusion techniques aim to improve upon the results of quantitative DWI. This meta-analysis compares the diagnostic performance of these advanced diffusion techniques including DWI, DTI and IVIM to assess whether they achieve an improvement in diagnostic performance that justifies their higher computational complexity and longer scan time which are needed to acquire the range of b-values or diffusion directions. Due to the lack of standardization in diffusion imaging, a sub-analysis investigates how acquisition sequence variations affect diagnostic performance.

3.2 Methods

3.2.1 Literature search

A search of PubMed and EMBASE was performed for studies involving women over the age of 18 between January 2000 and March 2018. The search terms for ADC studies included: *breast, diffusion, apparent diffusion coefficient, ADC, and mono-exponential*. The search terms for DTI studies included: *breast, diffusion tensor imaging, and DTI*. The search terms for IVIM studies included: *breast, diffusion, intra-voxel incoherent motion, IVIM, bi-exponential, and non-mono-exponential*. A search of the lists of references from included studies was also performed.

3.2.2 Study selection

Studies were included if they met the following eligibility criteria: (1) published in a peer reviewed journal (abstracts and conference proceedings excluded) (2) in English; (3) data obtained using a 1.5T or 3T MRI scanner, with MRI acquisition information reported; (4) DWI performed and ADC, DTI or IVIM parameters calculated; (5) the study's purpose was for investigating the diagnostic performance of ADC, DTI or IVIM, with criteria for classifying benign and malignant lesions clearly stated (i.e. threshold value used for a parameter; a computational method used), (6) Sufficient information was reported to extract the number of true positives (TP), false negatives (FN), false positives (FP) and true negatives

(TN) classified using the diagnostic criteria. If the values reported could not be reproduced, the study was excluded; (7) No limit was defined for age or sample size.

3.2.3 Data extraction

A data extraction spreadsheet was developed. The number of TP, FN, FP and TN using parameters ADC, MD, λ_1 , $\lambda_1 - \lambda_3$, FA, D, f, and D* were extracted. For studies that reported multiple sensitivities and specificities, the method that achieved the highest number of correctly classified lesions (TP + TN) to avoid overrepresentation of a sample was extracted. For studies that used both a training set and a test set, the test set values were extracted. Reader 1 was extracted when studies had multiple readers. For authors with multiple studies published in the same year, values from only one of the studies were extracted.

Other information extracted included: mean age and range or standard deviation (whichever reported), study design, MRI scanner and vendor, breast coil used, b-values (s/mm^2), repetition time (TR, msec), echo time (TE, msec), matrix size, slice thickness (mm), field-of-view (FOV, mm), parallel acceleration (GRAPPA/SENSE and acceleration factor), and fat suppression method. Corresponding authors were contacted for missing information.

3.2.4 Data quality assessment

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) was performed to assess the quality of studies and likelihood of bias [179]. Risk of bias was assessed in four domains: patient selection, methodology of the index test (parameter measurement and use of appropriate threshold to classify lesions), use of histology as a reference standard, and flow and timing. Funnel plots to visually examine publication bias were constructed. An asymmetric or skewed funnel plot suggested the presence of a publication bias. Asymmetry was quantified using Egger's test [180] with a p-value of < 0.05 indicating publication bias. The degree of heterogeneity between studies was measured by Cochran's Q-test and Higgins' I^2 test [181] using Meta-DiSc version 1.4 (Clinical Biostatistics Unit, Hospital Ramon y Cajal, Madrid, Spain; http://www.hrc.es/investigacion/metadisc_en.htm). A p-value < 0.05 for Cochran's Q-test or an I^2 value of $> 50\%$ indicated significant heterogeneity.

3.2.5 Statistical analysis

For each of the parameters (ADC, MD, λ_1 , $\lambda_1 - \lambda_3$, FA, D, f, and D*), forest plots for sensitivity and specificity were constructed. The bivariate model by Reitsma et al. [182] was used to estimate pooled sensitivities, specificities and areas under the curve (AUC) for all parameters

and construct summary receiver operating characteristics (sROC) curves. Analysis was carried out in R (version 3.1.3, R Foundation for Statistical Computing, Vienna, Austria) using the mada package.

Sensitivities, specificities and AUCs of studies evaluating ADC were compared to study the effect of b-value (minimum, maximum and number used), as well as other MRI parameters such as field strength, Vendor, use of partial field-of-view, slice thickness and resolution on diagnostic performance using a Student's T-test, Mann-Whitney U test or a one-way analysis of variance (ANOVA). Method of region of interest (ROI) delineation (use of the whole lesion, a small ROI or a single slice) was also compared with diagnostic performance using an ANOVA. A p-value < 0.05 indicated a statistically significant difference. The data was analysed in R.

3.3 Results

3.3.1 Study selection and data extraction

Using the key words, a search of the PubMed and EMBASE databases returned 515 ADC studies with 65 meeting the eligibility criteria. A search for DTI returned 71 studies with 6 meeting the eligibility criteria. A search for IVIM returned 80 studies with 9 meeting the eligibility criteria. 413 studies were excluded after a review of the titles and abstracts. The full text of the remaining 253 studies was reviewed and 173 excluded that did not meet the eligibility criteria. A total of 80 studies were included in the meta-analysis [39, 40, 44, 56, 58–62, 183–203, 176, 204–245]. Six studies evaluated both ADC and IVIM and one study evaluated both ADC and DTI for all patients included in the study. Figure 3.1 shows a flow diagram of article exclusion. Details of included studies are provided in Appendix 1. 6791 lesions (3930 malignant and 2861 benign) from 73 eligible studies were included. There was a large range of reported mean ADC values of malignant ($0.66\text{--}1.50 \times 10^{-3} \text{ mm}^2/\text{s}$) and benign lesions ($0.87\text{--}2.00 \times 10^{-3} \text{ mm}^2/\text{s}$). Reported diagnostic threshold ADC values ranged from $0.87 \times 10^{-3} \text{ mm}^2/\text{s}$ to $2 \times 10^{-3} \text{ mm}^2/\text{s}$. A number of studies using DTI and IVIM used a combined thresholds approach. The sensitivities and specificities of these studies are compiled in Table 3.1. Jiang et al. reported a sensitivity and specificity for D^* and f combined [246] whereas Bokacheva et al. reported a combination of D and f using linear discriminant analysis [216]. Dijkstra et al. used all 3 IVIM parameters [188] and Lima et al. reported a combination of f , ADC and diffusion kurtosis coefficient K [247]. Jiang et al. also reported a combination of fractional anisotropy and the maximal anisotropy index [59].

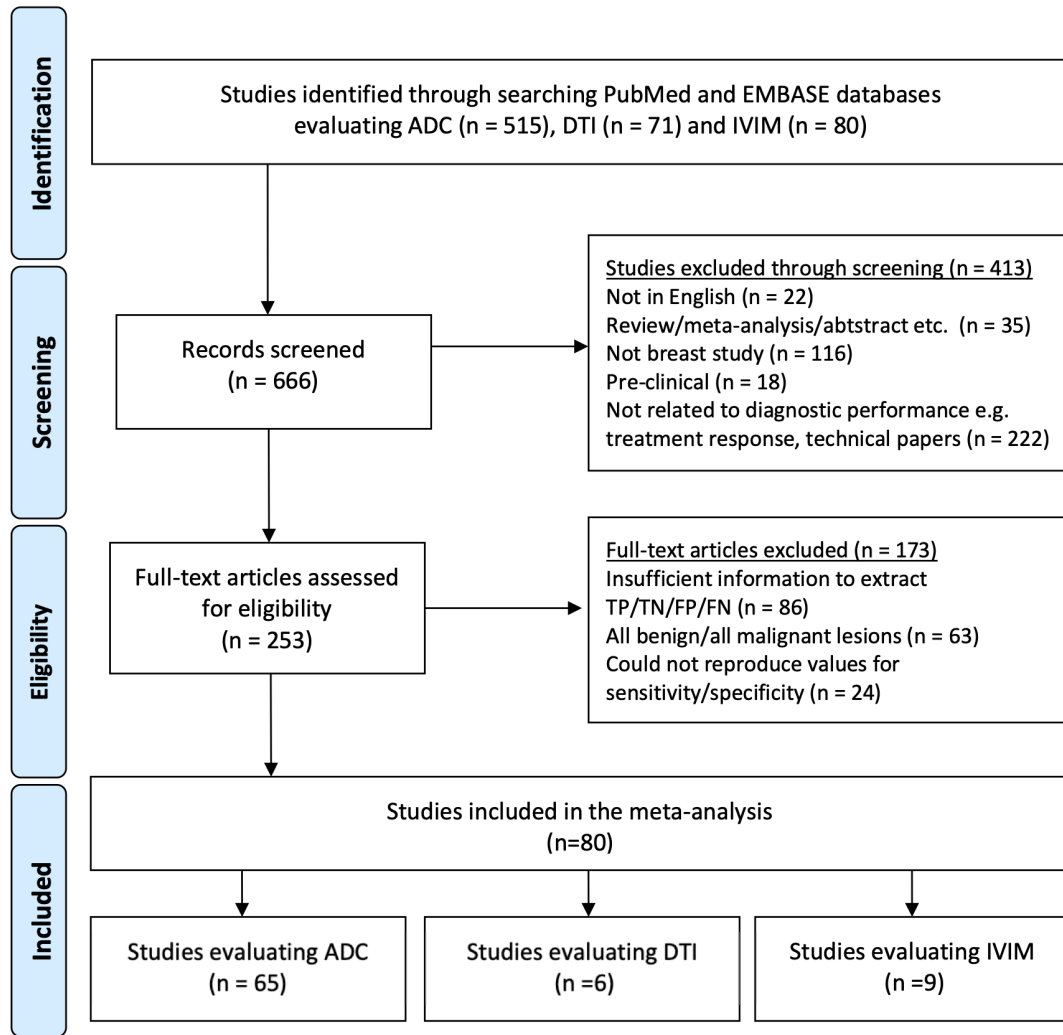


Figure. 3.1 Flow diagram for selection and exclusion of studies.

Table 3.1 Combined approaches using DTI and IVIM parameters

Author	Sensitivity	Specificity	Method
Jiang et al. [246]	80.6%	74.3%	Combined thresholds - D*, f
Iima et al. [247]	94.7%	75.0%	Combined thresholds - f, ADC, K
Jiang et al. [59]	85.3%	90.9%	Combined thresholds – FA and $\lambda_1 - \lambda_3$
Dijkstra et al. [188]	92.2%	52.2%	Combined thresholds - D, D* and f
Bokacheva et al. [216]	85.0%	86.0%	Linear discriminant analysis (LDA) - D, f

D*: pseudo-diffusion, f: perfusion fraction, ADC: apparent diffusion coefficient, K: diffusion kurtosis coefficient, FA: fractional anisotropy, $\lambda_1 - \lambda_3$: maximal anisotropy index, D: tissue diffusivity

3.3.2 Data quality assessment

Figure 3.2 shows the distribution of QUADAS-2 scores for risk of bias. The majority of studies had a low risk of bias. Some studies were marked as unclear concerning patient selection due to missing inclusion or exclusion criteria. Common weaknesses included the lack of justification for diagnostic threshold ($n=19$), where receiver operating characteristic curves or other analyses were not used, or the lack of consistent use of histology as the reference standard for all patients ($n=11$).

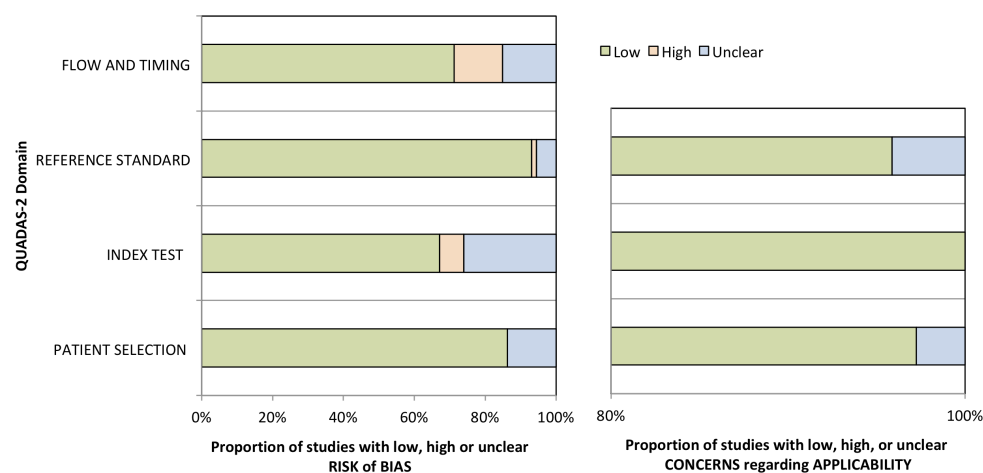


Figure. 3.2 Results of quality assessment for risk of bias and concerns regarding applicability of included studies. Quality assessment of diagnostic accuracy studies (QUADAS-2) scores for each category are expressed by percentages of studies that have a low, high or unclear risk of bias.

The funnel plot for ADC studies has a funnel-shaped distribution and the lack of studies in the bottom left quadrant indicates publication bias (Figure 3.3). Significant asymmetry was measured using Egger's test ($p < 0.001$). Due to the low number of studies for each of the other parameters, funnel asymmetry and Egger's test were not assessed.

3.3.3 Statistical analysis

Table 3.2 shows the results of the pooled analysis of ADC, DTI, and IVIM parameters. High heterogeneity between studies was measured for most parameters except f . I^2 values of 0% were measured for the specificities of λ_1 and FA, however the low number of studies included in the analysis ($n=3$ for both) resulted in an undefined I^2 value as opposed to significant lack

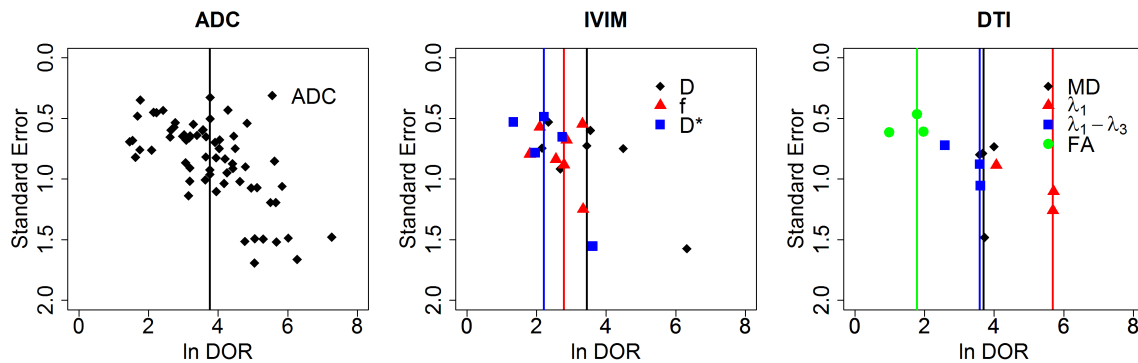


Figure. 3.3 Funnel plots for apparent diffusion coefficient (ADC), intra-voxel incoherent motion (IVIM) and diffusion tensor imaging (DTI) parameters. Log of diagnostic odds ratio (DOR) is plotted against standard error (SE). The vertical line represents the median. IVIM parameters include tissue diffusivity (D), perfusion fraction (f) and pseudo-diffusion (D^*). DTI parameters include mean diffusivity (MD), the prime diffusion coefficient (λ_1), the maximal anisotropy index ($\lambda_1 - \lambda_3$) and the fractional anisotropy (FA).

of heterogeneity as Cochran's p-values were measured as 0.59 and 0.61, respectively. The pooled AUC of the apparent diffusion coefficient (ADC) was 0.92. The highest performing parameter for DTI was the prime diffusion coefficient (λ_1) with a pooled AUC of 0.94. The highest performing parameter for IVIM was tissue diffusivity (D) with a pooled AUC of 0.90. A forest plot for sensitivity and specificity is presented for all 3 parameters in Figure 3.4. sROC curves are presented in Figure 3.5.

Table 3.2 Results of pooled estimates and heterogeneity measures for diffusion MRI studies of the breast

Parameter	Studies	Lesions (Patients)	Sensitivity (%) (95% CI)	Heterogeneity		Specificity (%) (95% CI)	Heterogeneity		AUC (95% CI)
				Cochran's Q p-value	I ² (%)		Cochran's Q p-value	I ² (%)	
ADC	65	6408 (5892)	89 (87 - 91)	<0.001	71	82 (78 - 85)	<0.001	86	0.92 (0.91 - 0.93)
D	7	536 (486)	88 (79 - 92)	0.001	72	79 (64 - 89)	<0.001	81	0.90 (0.85 - 0.96)
f	7	397 (366)	81 (74 - 86)	0.34	12	76 (64 - 85)	0.06	50	0.85 (0.81 - 0.90)
D*	5	334 (309)	82 (67 - 91)	<0.001	87	61 (37 - 80)	<0.001	85	0.80 (0.74 - 0.87)
λ_1	3	201 (181)	93 (80 - 98)	0.10	56	90 (81 - 95)	0.59	0	0.94 (0.91 - 0.96)
MD	4	262 (247)	90 (79 - 96)	0.03	66	78 (51 - 92)	0.001	86	0.92 (0.90 - 0.97)
$\lambda_1 - \lambda_3$	3	201 (181)	73 (36 - 93)	<0.001	92	89 (62 - 97)	0.004	82	0.89 (0.77 - 1.00)
FA	3	219 (200)	64 (42 - 81)	0.003	83	74 (62 - 84)	0.61	0	0.76 (0.68 - 0.87)

ADC: apparent diffusion coefficient, D: tissue diffusivity, f: perfusion fraction, D*: pseudo-diffusion, λ_1 : prime diffusion coefficient, MD: mean diffusivity, $\lambda_1 - \lambda_3$: maximal anisotropy index, FA: fractional anisotropy, CI: confidence interval, AUC: area under the receiver operating characteristic curve

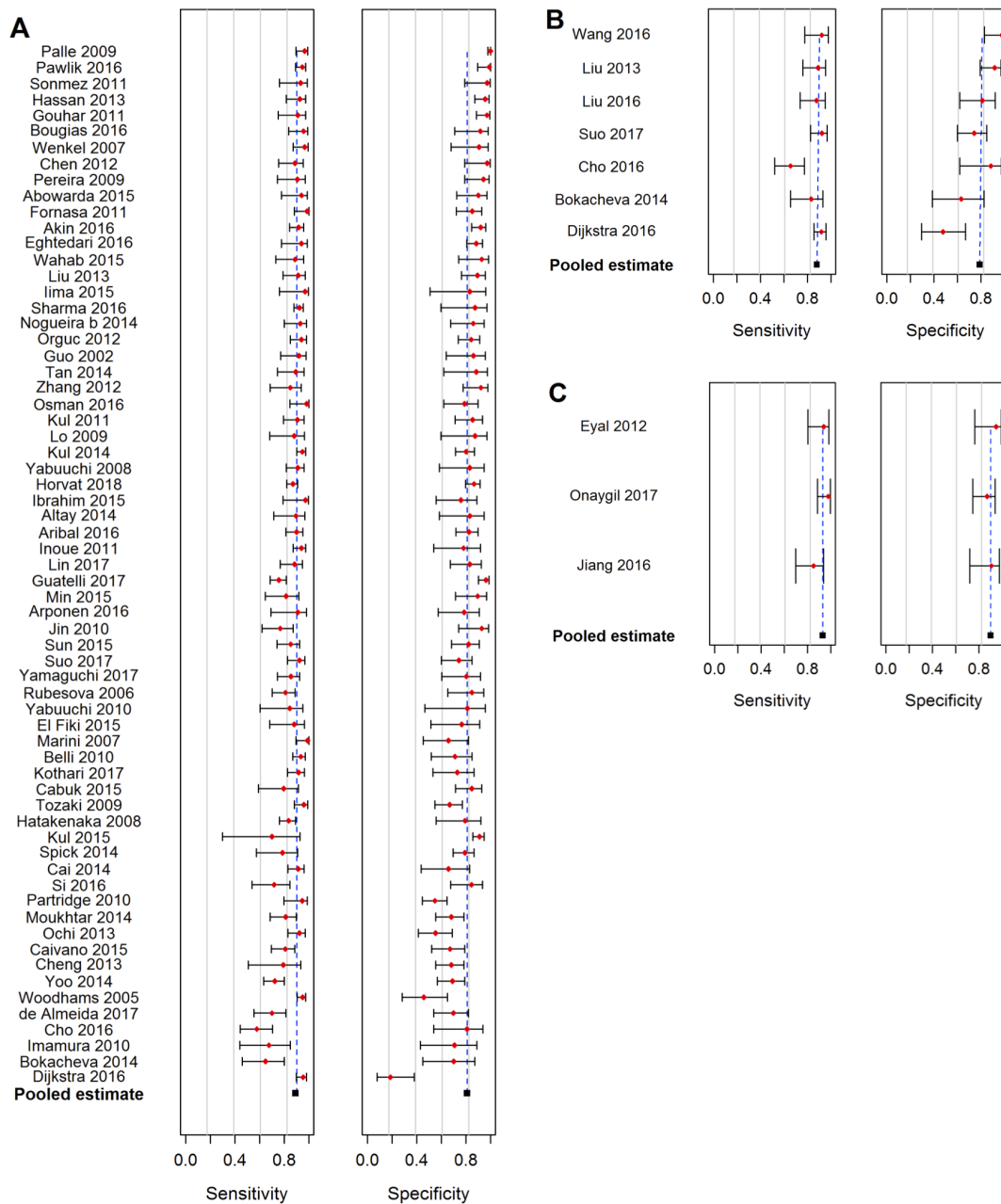


Figure. 3.4 Forest plot of sensitivity and specificity with 95% confidence intervals using A) the apparent diffusion coefficient (ADC), B) tissue diffusivity (D) and C) the prime diffusion coefficient (λ_1), ordered from high to low sensitivity and specificity. Vertical lines denote pooled summary estimates of sensitivity and specificity.

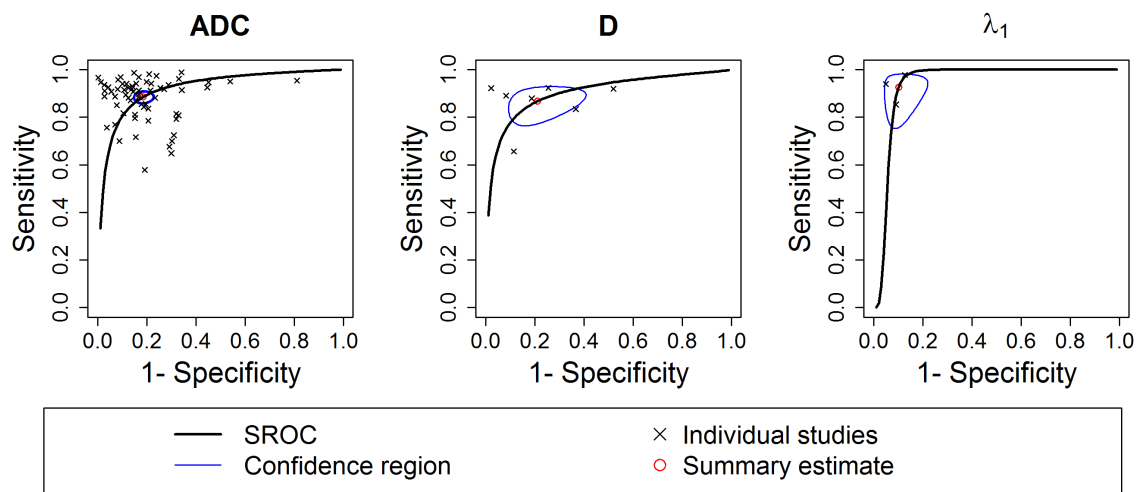


Figure 3.5 Summary receiver operating characteristics (sROC) curves using the bivariate model with 95% confidence regions. The pooled area under the curve (AUC) was 0.92 for the apparent diffusion coefficient (ADC), 0.90 for tissue diffusivity (D) and 0.94 for the prime diffusion coefficient (λ_1).

Table 3.3 shows the results of the sub-analysis. For studies using the ADC, choice of minimum b-value of 0 s/mm² (n=56) or 50 s/mm² (n = 9) had no significant effect on sensitivity or specificity (p = 0.82 and p = 0.52, respectively). No significant differences (p > 0.05) were found in sensitivity and specificity for maximum b-value, number of b-values, field strength, vendor, partial field-of view, slice thickness, resolution or method of ROI delineation.

Table 3.3 Results of sub-analysis comparing studies using the ADC .

	Sensitivity p-value	Specificity p-value
Minimum b-value (0 or 50 s/mm ²)	0.82	0.52
Maximum b-value	0.08	0.71
Number of b-values	0.84	0.94
Field strength (1.5T or 3T)	0.14	0.64
Vendor	0.93	0.78
Partial field-of-view	0.79	0.43
Slice thickness	0.60	0.72
Resolution	0.90	0.65
Method of ROI delineation	0.66	0.57

3.4 Discussion

While other meta-analyses have assessed the diagnostic performance of the ADC model [38, 177, 178], this is the first study to systematically compare all relevant advanced non-Gaussian diffusion techniques with standard DWI for quantitatively distinguishing benign and malignant lesions. The pooled estimates of sensitivity, specificity and AUC were found to be comparable for ADC, IVIM and DTI. However, due to the low number of studies included and large confidence intervals, this meta-analysis lacks the statistical power to conclude that they are diagnostically equivalent.

The pooled sensitivities and specificities using DWI, IVIM and DTI in this meta-analysis were comparable to the pooled sensitivity and specificity of dynamic contrast-enhanced MRI (93% and 71%, respectively)[10]. Other non-Gaussian diffusion models have been proposed, such as diffusion kurtosis and the stretched exponential model, though these were not investigated due to a low number of publications to date. A study by Suo et al. found that kurtosis and stretched exponential achieved a better goodness-of-fit, though the AUCs for non-mono-exponential models were comparable to the ADC, in accordance with these findings [56].

DTI lacks standardization in methodology and reporting of parameters. The prime diffusion direction λ_1 and the mean diffusivity achieved a diagnostic accuracy equal to or greater than the ADC; however, the number of eligible studies included is very low. While it is suggested that reduced structuring in malignant breast lesions should be reflected by a reduced diffusion anisotropy [61], anisotropy measures achieved a mixed diagnostic utility, with some studies finding no significant difference in fractional anisotropy between malignant and benign lesions [58, 60]. The number of diffusion directions used ranges from 6 to 64, though the b-value pair of 0 and 1000 s/mm² was the most commonly used [58–61].

While the increasingly used technique of IVIM in the breast achieves a high diagnostic accuracy, there is also still a lack of consistent methodology. There is a large variation in the number and range of b-values used and in the choice of parameters reported, with studies often using a combined thresholds approach. Variations in MRI technique prevent determination of generalized threshold values as ADC quantification depends on the choice of b-values [248]. A number of single-center studies have reported their optimal b-value combination. Bogner et al. reported optimal ADC determination and DWI quality at b = 50 and 850 s/mm² [249]. Dorrius et al. found that b = 0 and 1000 s/mm² were the optimum values, finding that this b-value combination achieved the highest percentage difference in ADC of benign and malignant lesions [250]; this was the most commonly used pair of b-values in this meta-analysis (n = 29). It has been suggested that using more b-values achieves a better separation of diffusion and perfusion [251], particularly at low b-values

where the contribution of perfusion to signal decay is strongest [49]. While this may suggest avoiding low b -values, the precise b -value threshold for minimising perfusion effects has not been standardized and choice of minimum b -value showed no significant difference in diagnostic performance in studies in this meta-analysis, though this may be due to the lack of statistical power. Also, the diagnostic accuracy of D , corresponding to an ADC measurement with effects of perfusion excluded, was comparable to standard DWI. While there is no consensus on whether excluding $b = 0$ or avoiding low b -values constitutes ‘excluding perfusion effects’, both have been shown to have limited effect on diagnostic performance. Choice of fat suppression technique, such as short tau inversion recovery (STIR) or spectral adiabatic inversion recovery (SPAIR), has been shown to influence image quality and ADC quantification, though diagnostic performance was comparable [252]. These discrepancies highlight the importance of choosing similar protocols and methods of data analysis to compare studies across multiple centers.

This meta-analysis has some limitations. First, the low number of studies contributing to the pooled estimates resulted in large confidence intervals, limiting the conclusions that can be drawn from the comparable AUCs. Second, overrepresentation of a sample may be a limitation of pooled estimates since multiple studies were included from the same author that may have used the same patient population. Third, for studies that did not report TP, FN, FP and TN, these outcomes were calculated from sensitivity, specificity, and number of malignant and benign lesions. However, many studies were excluded ($n=24$) because they resulted in a non-integer number of lesions. Finally, due to low numbers of publications, other non-mono-exponential models were not included.

In conclusion, diffusion-weighted imaging (DWI), diffusion tensor imaging (DTI), and intra-voxel incoherent Motion (IVIM) are able to discriminate between malignant and benign lesions with a high sensitivity and specificity. IVIM is diagnostically comparable to apparent diffusion coefficient (ADC), though the exclusion of perfusion effects using the tissue diffusion coefficient, D , does not improve upon the results of the ADC model. DTI achieves a higher accuracy than the ADC, though the number of studies to date is limited. Both IVIM and DTI lack standardization in methodology and parameters reported.

Chapter 4

Improving the image quality of diffusion-weighted imaging (DWI) in breast cancer

In this chapter, a DWI protocol using multiplexed sensitivity encoding (MUSE), a multi-shot DWI technique, is optimised through a series of phantom experiments and validated in 20 patients. This acquisition technique aims to overcome some of the limitations of conventional DWI by reducing blurring and distortion. Significant improvements in qualitative and quantitative image quality metrics are demonstrated. As a result of this work, this protocol has been adopted as part of the routine clinical breast imaging protocol at Addenbrooke's hospital.

Contents of this chapter have been published in the *British Journal of Radiology* [253] and presented at the International Society for Magnetic Resonance in Medicine (ISMRM) 2019 and 2020 (abstracts #4121 and #2331).

4.1 Introduction

Diffusion-weighted imaging (DWI) is a promising non-contrast MRI technique in the detection of breast cancer [51, 67]. However, diffusion-weighted images acquired using single-shot EPI (ss-EPI) are limited by image artefacts and suffer from blurring due to T_2^* decay during readout. The relatively low spatial resolution results in the averaging of tumour volumes with nearby breast tissues, particularly for small lesions and non-mass enhancements which are prone to volume-averaging affects due to the large voxel sizes of DWI compared to dynamic

contrast-enhanced MRI (DCE-MRI) [254]. In order for diffusion MRI to be clinically relevant, it must be able to detect and characterise all lesions, even small lesions.

Parallel imaging techniques (such as SENSE [166]) aim to reduce artefacts in images acquired using EPI by reducing the length of the echo train and therefore reducing the T_2^* blurring. However, the extent to which parallel imaging can improve image quality is limited by the coil hardware. Multi-shot techniques also attempt to reduce the length of the echo train by acquiring k -space in a number of segments. These techniques must account for motion between shots, often using a ‘navigator’ pulse to acquire a low-resolution image that can be used to correct for motion-induced phase errors between shots.

Multiplexed sensitivity encoding (MUSE), a multi-shot segmented EPI technique, expands on existing sensitivity-encoding techniques by acquiring k -space with an interleaved trajectory with the aim of achieving better spatial resolution and high signal-to-noise ratio (SNR) without the need for navigator pulses between interleaves [175]. Benefits of a navigator-free method include an increase in efficiency of acquisition as navigators require 30-40ms per excitation [174]. Similar to parallel acceleration, MUSE is limited by the design of the receiver coil and the number of coil components. While MUSE has not yet been investigated in the breast, a number of studies have investigated the improvement in image quality and diagnostic performance of RESOLVE, a readout-segmented EPI technique [58, 254–257].

This study aimed to compare the image quality of DWI acquired using ss-EPI and MUSE, and investigate the effect of MUSE, used in conjunction with parallel acceleration, on the quantification of the apparent diffusion coefficient (ADC) in a breast phantom and in malignant, benign and normal breast tissue in a small cohort of patients. In addition, an optimised parallel acceleration factor and number of shots for MUSE-DWI was determined.

4.2 Methods

4.2.1 Phantom study design

The effect of varying parallel acceleration factor and number of shots was first evaluated in a phantom. The phantom used in this study was manufactured by QMD (High Precision Devices, Inc., Boulder, CO), and consisted of two silicon breast inserts designed to investigate diffusion, distortion and T_1 relaxation. The diffusion insert contained 16 vials of solutions arranged in two concentric rings with diameter and length 15mm x 110mm (inner ring) and 15mm x 80mm (outer ring) (Figure 4.1). 12 vials contained varying concentrations of polyvinylpyrrolidone (PVP) in water (0, 10, 14, 18, 25, and 40%) to mimic the apparent diffusion coefficients of benign and malignant tumours found in the literature and 4 vials

contained a fat mimic. The interstitial space was filled with a solution of 35% corn syrup in water to mimic fibroglandular tissue. The phantom was kept in the scanner room at room temperature (18°C).

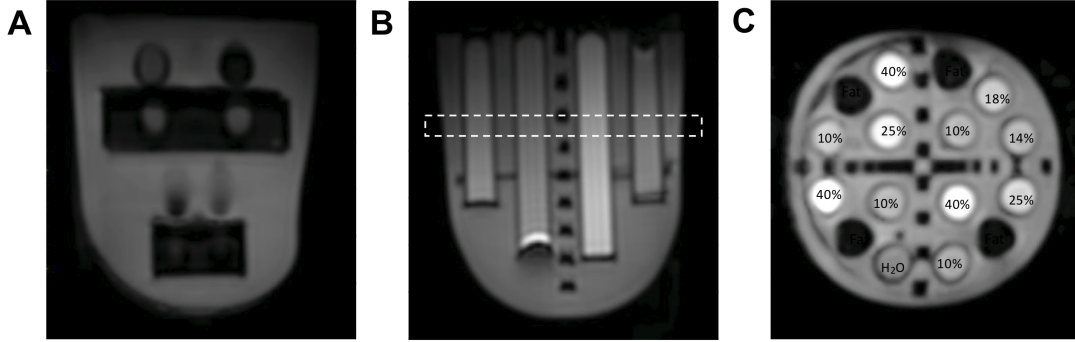


Figure. 4.1 Axial MUSE-DWI (2 shots, acceleration factor 1.5) images of the A) T_1 relaxation and B) diffusion phantom inserts. The dotted line refers to the coronal slice C) with contents of the 16 vials labelled. Percentages refer to the percentage of polyvinylpyrrolidone (PVP) in water.

All phantom acquisitions were performed on a 3T system (Discovery MR 750, GE Healthcare, Waukesha, WI, USA) using a dedicated bilateral eight-channel phased-array breast coil. DWI was performed using ss-EPI and 6 permutations of MUSE-DWI with varying parallel acceleration factors and numbers of shots (see Table 4.1). A T_1 -weighted image (TR/TE = 659/15.3ms, matrix = 512x512, slice thickness = 3.5mm) was acquired using a fast spin echo (FSE) sequence as a geometric reference.

A measurement of blurring was obtained using the Cr  t  -Roff  t blur metric [258], previously adapted from the field of computer vision [259], for a central slice through the phantom for all acquisitions. The metric is quantified by comparing the perceptible levels of blurring between the input image and a version of the image which has been further blurred using a low-pass filter. Values of the blur metric range from 0 (sharp) to 1 (blurry). A measurement of distortion was obtained by calculating the Mattes Mutual Information (MI) metric [260] between the $b = 800 \text{ s/mm}^2$ and T_1W images, resampled to the same matrix size as the DWI images. Values of the MI metric range from 0 (distorted) to 1 (not distorted). Metrics were calculated for each slice and averaged over the whole image volume.

For each of the 16 vials, regions of interest (ROIs) were drawn on axial images for each of the diffusion gels using Osirix (version 8.0.1, Pixmeo, Switzerland). Each ROI consisted of 3 rectangles drawn on contiguous slices, avoiding the edges of the vials to minimise partial volume effects. Generation of ADC maps and diffusion analyses were carried out using

in-house software developed in MATLAB (The Mathworks, version 2019a). The mean ADC was measured for each ROI and the coefficient of variation calculated between all acquisitions.

Table 4.1 Scan parameters

	Phantom studies		Patient studies		
	ss-EPI-DWI	MUSE-DWI	ss-EPI-DWI	MUSE-DWI	DCE-MRI
Slice Orientation	Axial	Axial	Axial	Axial	Axial
b-values (s/mm ²)	0, 800	0, 800	0, 800	0, 800	-
TR/TE (ms)	4000/ 51	4000/ 54-55	3714-14853/ 60.7	8203-16715/ 71.3	5.1/ 2.4
Matrix size	160 x 160	160 x 160	96 x 128	160 x 160	384 x 384
Field-of-view (mm ²)	360 x 360	360 x 360	360 x 360	360 x 360	360 x 360
Slice thickness (mm)	3.5	3.5	3.5	3.5	2.0
Acceleration factor	2	1, 1.5, 2	2	1.5	2.5
Number of shots	1	2, 3	1	2	-
Acquisition Time	1 min 8s	1 min 52s, 2 min 48s	6 – 8 min	8 min	9 min

ss-EPI: single-shot echo-planar imaging, DWI: diffusion-weighted imaging, MUSE: multiplexed sensitivity encoding, DCE: dynamic contrast-enhanced, TR: repetition time, TE: echo time.

4.2.2 Patient study design

Between October 2018 and October 2019, 20 women were scanned under an existing research ethics protocol. The local institutional review boards and ethics committees approved this study and written informed consent was obtained from all patients. Patients were recruited if they had pathologically confirmed invasive breast cancer and had no previous treatment (such as breast surgery or neoadjuvant chemotherapy).

All clinical acquisitions were performed on a 1.5T system (MR 450W, GE Healthcare, Waukesha, WI) using an 8-channel breast coil. ss-EPI-DWI and MUSE-DWI were performed.

High resolution T₁W dynamic contrast enhanced (DCE) images were also acquired. Scan parameters are given in Table 4.1.

4.2.3 Image analysis

Qualitative analysis

Two breast radiologists, each of whom had at least 5 years of experience, independently reviewed all the images. Both ss-EPI and MUSE images were scored based on three qualitative image criteria: lesion conspicuity, contrast between lesion and tissue and diagnostic confidence. All criteria were scored from 1 (poor) to 5 (excellent). Lesions were also assessed for multifocality and multi-centricity. The readers were not blinded to the scan type and images were read consecutively.

Quantitative analysis

The Cr  t  -Roffet blur metric was calculated for ss-EPI and MUSE images, and the MI distortion metric was calculated between ss-EPI and MUSE images and the corresponding DCE images, resampled to the same matrix size as the DWI images. Metrics were calculated for each slice and averaged over the whole image volume.

ADC maps were generated using in-house software developed in MATLAB (version 2019a). Regions of interest were manually drawn on the ss-EPI and MUSE b₈₀₀ images by two breast radiologists in consensus with reference to DCE images to aid tumour delineation. As previously described by Wisner et al. [254], to obtain a measurement of signal in normal tissue the ROI for each lesion was copied and placed on contralateral fibroglandular tissue. These ROIs were copied to the ADC maps and mean ADC values were measured for each lesion (ADC_{lesion}) and normal fibroglandular tissue (ADC_{tissue}) ROI. To account for the difference in acquired voxel size, the normalised ADC (nADC) was calculated using $nADC = ADC_{lesion}/ADC_{tissue}$.

Statistical analysis

Qualitative and quantitative image quality metrics for images using ss-EPI or MUSE were compared on a per-patient basis. ROI sizes and ADC values were compared on a per-lesion basis. All comparisons used the paired t-test or a Wilcoxon signed-rank test. The inter-observer agreement of the qualitative scoring criteria was assessed using Cohen's Kappa. $P < 0.05$ was considered statistically significant for all statistical tests. All statistical analyses

were carried out using R (R versions 3.1.3; R Foundation for Statistical Computing, Vienna, Austria).

4.3 Results

4.3.1 Phantom results

Images acquired using ss-EPI and all permutations of MUSE are shown in Figure 4.2. ADC values measured using the phantom were consistent with values from the literature [261]. The variation in ADC for each acquisition is shown in Table 4.2. There was a low coefficient of variation for measured ADC value ($<2\%$) for all PVP concentrations. The variation in MI metric and Cr  t  -Roffet blur metric between acquisitions is shown in Figure 4.3. Both distortion and blurring improved (an increase in the MI metric and a decrease in the Cr  t  -Roffet blur metric) with an increasing number of shots and acceleration factor. Increasing the acceleration factor when using 3 shots did not improve distortion. For clinical implementation, an acceleration factor of 1.5 and 2 shots were chosen as a pragmatic compromise between acquisition time and image quality.

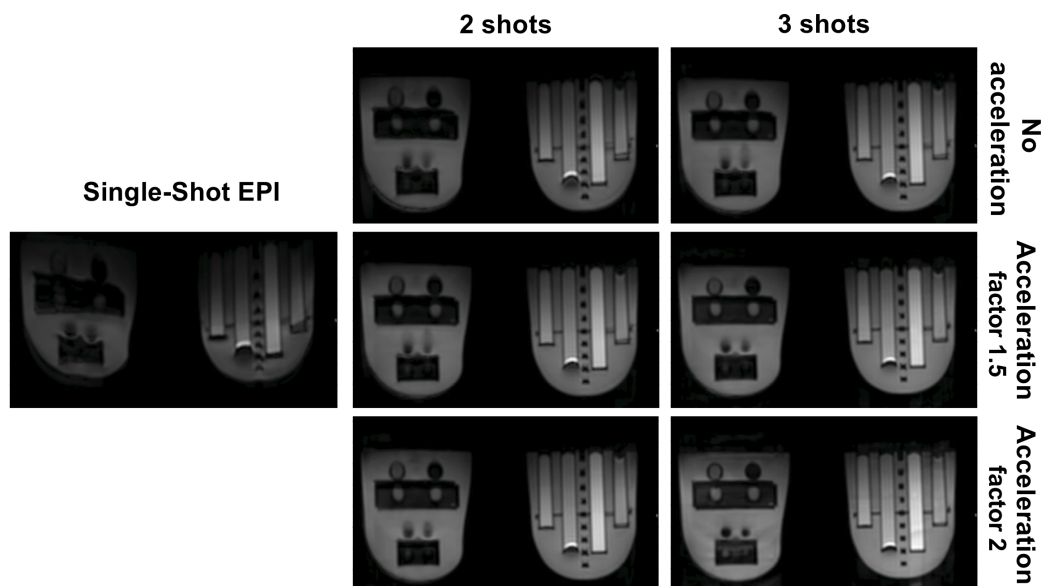


Figure. 4.2 Axial images of the breast phantom using single-shot echo-planar imaging (ss-EPI), 2 or 3 shots and acceleration factors of 1, 1.5 and 2. Insertion of the phantom in the breast coil causes the left phantom insert to bend slightly- this distortion is not caused by imaging technique.

Table 4.2 Apparent Diffusion Coefficient values measured for each acquisition for varying PVP concentrations.

PVP concentration (%)	Apparent diffusion coefficient (10 ⁻³ mm ² /s)								Coefficient of variation (%)
	Manufacturer	ss-EPI	MUSE 2 shots			MUSE 3 shots			
			Acceleration Factor			Acceleration Factor			
			1	1.5	2	1	1.5	2	
0	2.02	2.40	2.39	2.43	2.41	2.42	2.43	2.41	1.2
10	1.66	1.74	1.75	1.79	1.79	1.79	1.81	1.82	1.5
14	1.50	1.58	1.48	1.48	1.48	1.49	1.50	1.51	1.9
18	1.34	1.37	1.29	1.31	1.31	1.31	1.32	1.33	1.8
25	1.08	1.19	1.09	1.10	1.11	1.11	1.12	1.13	1.6
40	0.68	0.67	0.60	0.70	0.69	0.69	0.66	0.67	1.6

PVP: polyvinylpyrrolidone, ss-EPI: single-shot echo-planar imaging, MUSE: multiplexed sensitivity encoding

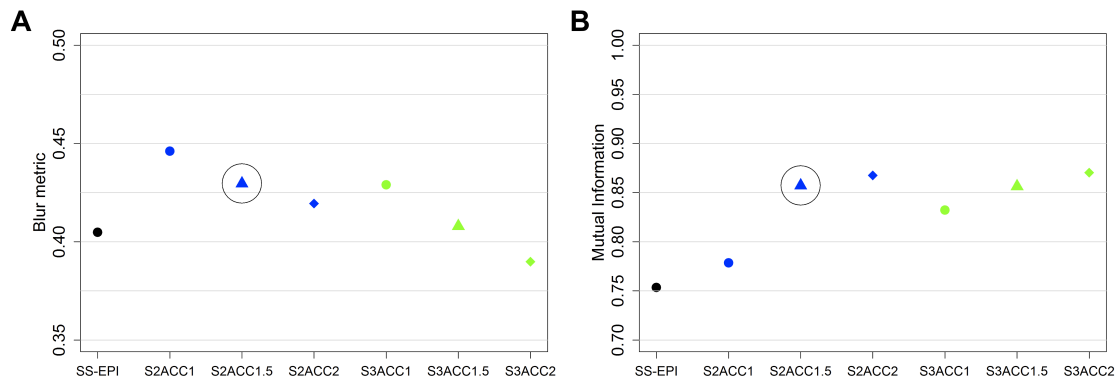


Figure. 4.3 Values of A) the Crété-Roffet blur metric and B) the mutual information metric measured for each phantom acquisition. S1, S2 and S3 refer to the use of 1, 2 and 3 shots. ACC1, ACC1.5 and ACC2 refer to the use of acceleration factor 1, 1.5 and 2. Circles indicate the acceleration factor and number of shots used in the clinical protocol.

4.3.2 Patient results

20 patients (median age 51 years, range 28 – 81 years) were imaged. A total of 22 malignant lesions and 3 benign lesions were identified (median size 16mm, range 10 – 56mm). There

were 5 patients that had more than one lesion. 2 patients had a second lesion located in the contralateral breast. 3 patients had a second lesion located in the same breast. Details of the patient and lesion characteristics are given in Table 4.3. Comparisons of image quality for an invasive ductal carcinoma, an invasive lobular carcinoma, and a patient with an invasive carcinoma with mixed ductal and lobular growth patterns and columnar and fibrocystic changes with benign calcification are shown in Figure 4.4.

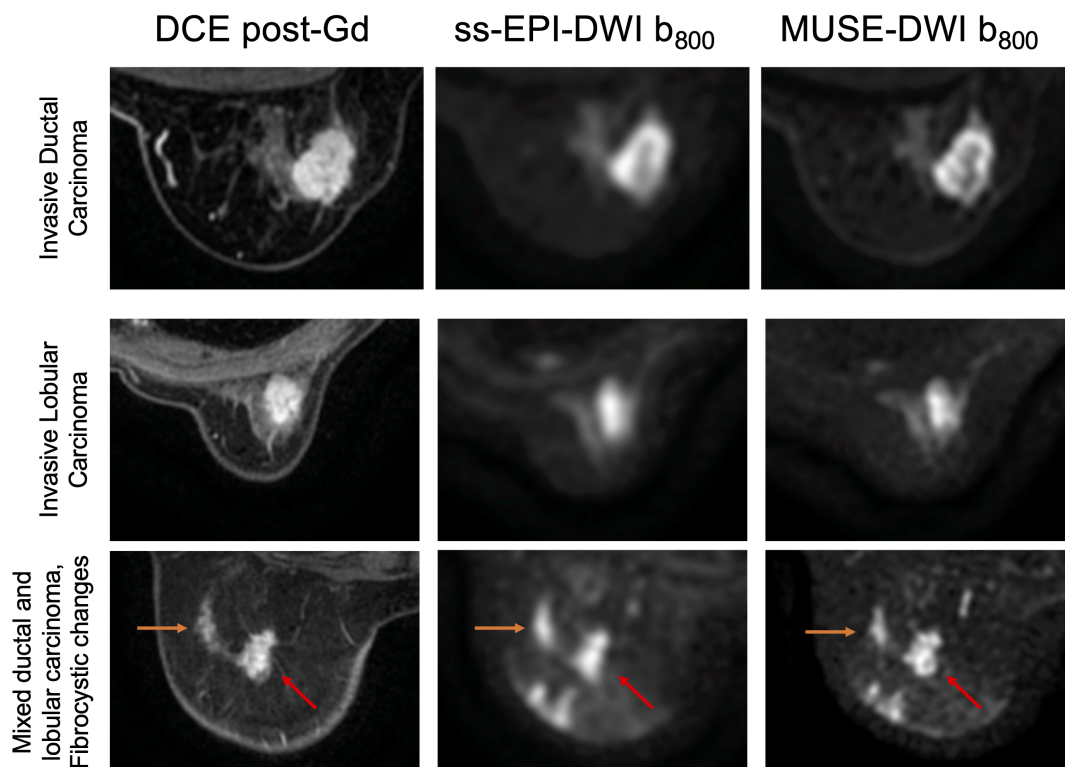


Figure. 4.4 Post-gadolinium (Gd) DCE T₁W images, DWI using ss-EPI and DWI using MUSE for a 60 year old woman with invasive ductal carcinoma (top), a 63 year old woman with invasive lobular carcinoma (middle), and a 51 year old woman with an invasive carcinoma with mixed ductal and lobular growth patterns (red arrow) and columnar and fibrocystic changes with benign calcification (orange arrow) (bottom).

Table 4.3 Patient and lesion characteristics

Mean age in years \pm SD (range)	53.8 \pm 12.0 (28 - 81)
Total lesions (n = 24)	
Lesion size in mm (mean \pm SD)	19.58 \pm 11.68
Range	10 – 56
	N (%)
Malignant lesions (n = 22)	
Invasive ductal carcinoma	11 (50.0)
Invasive lobular carcinoma	8 (36.4)
Invasive carcinoma with mixed ductal and lobular features	1 (4.5)
Invasive ductal carcinoma with mucinous differentiation	1 (4.5)
High grade ductal carcinoma in situ	1 (4.5)
Benign Lesions (n = 3)	
Fibroadenoma	1 (33.3)
Fibrocystic change with columnar cell changes	1 (33.3)
Columnar and fibrocystic changes with benign calcification	1 (33.3)
Invasive breast cancer grade	
Grade 1	3 (14.3)
Grade 2	14 (66.7)
Grade 3	4 (19.0)
ER status	
Positive	20 (95.2)
Negative	1 (4.8)
PR status	
Positive	20 (95.2)
Negative	1 (4.8)
HER2 status	
Positive	1 (4.8)
Negative	20 (95.2)

SD: standard deviation, ER: estrogen receptor, PR: progesterone receptor, HER2: human epidermal growth factor receptor 2

4.3.3 Qualitative results

Results of qualitative comparisons are given in Table 4.4. MUSE was superior to ss-EPI for contrast between lesion and tissue, and significantly superior for lesion conspicuity and diagnostic confidence criteria. Inter-reader agreement as measured by Cohen's kappa was higher for the scoring of MUSE images compared to ss-EPI.

Table 4.4 Qualitative comparisons.

Criteria	Score range	ss-EPI	MUSE	p-value	κ ss-EPI	κ MUSE
Lesion conspicuity	1-5	3.5 ± 1.1	4.1 ± 0.9	<0.001	0.05	0.4
Lesion-tissue contrast	1-5	3.8 ± 1.2	3.9 ± 1.0	0.86	- 0.007	0.3
Diagnostic confidence	1-5	3.7 ± 1.1	4.2 ± 1.1	<0.001	0.08	0.2

ss-EPI: single-shot echo-planar imaging, MUSE: multiplexed sensitivity encoding, κ : Cohen's Kappa
p-values were determined using a Wilcoxon signed-rank test.

4.3.4 Quantitative results

Results of quantitative comparisons are given in Table 4.5. The distributions of blur and distortion metrics using ss-EPI and MUSE are shown in Figure 4.5. The Cr  t  -Roffet blur metric was significantly lower for MUSE-DWI than for ss-EPI-DWI ($p < 0.001$), indicating less blurring. The MI metric was significantly higher for MUSE-DWI than for ss-EPI-DWI ($p = 0.002$), indicating more similarity to DCE images and therefore less distortion.

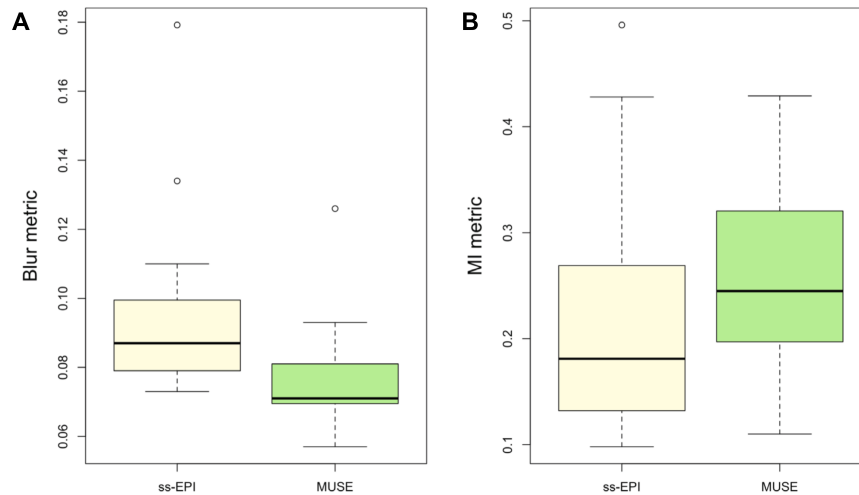
4.3.5 Comparison of ADC values: ss-EPI versus MUSE

The distributions of ADC values measured using ss-EPI and MUSE are shown in Figure 4.6A. The mean ADC of malignant lesions was $1.49 \pm 0.30 \text{ mm}^2/\text{s}$ using ss-EPI and $1.23 \pm 0.27 \text{ mm}^2/\text{s}$ using MUSE. The mean ADC of benign lesions was $1.73 \pm 0.27 \text{ mm}^2/\text{s}$ using ss-EPI and $1.42 \pm 0.25 \text{ mm}^2/\text{s}$ using MUSE. ADC values measured using MUSE were significantly lower than those measured using ss-EPI for malignant lesions but not benign lesions ($p < 0.001$ and $p = 0.21$, respectively). ADC values measured were in agreement with those in the literature [261], though the mean ADC for malignant lesions measured using ss-EPI was high.

Table 4.5 Quantitative comparisons

Parameter	ss-EPI	MUSE	p-value
Mean ADC ($\times 10^{-3} \text{ mm}^2/\text{s}$)			
Fibroglandular tissue	2.24 ± 0.45	1.76 ± 0.32	<0.001
Benign Lesions	1.73 ± 0.27	1.42 ± 0.25	0.21
Malignant lesions	1.49 ± 0.30	1.23 ± 0.27	<0.001
Mean normalised ADC			
Benign Lesions	0.79 ± 0.21	0.97 ± 0.23	0.28
Malignant lesions	0.69 ± 0.24	0.72 ± 0.18	0.62
Crété-Roffet blur metric	0.10 ± 0.02	0.08 ± 0.02	<0.001
Mattes Mutual Information metric	0.19 ± 0.12	0.23 ± 0.10	0.002

ss-EPI: single-shot echo-planar imaging, MUSE: multiplexed sensitivity encoding, ADC: apparent diffusion coefficient. p-values were determined using a paired t-test

**Figure. 4.5** Comparison of A) Crété-Roffet blur metrics and B) Mattes Mutual Information distortion metrics using ss-EPI and MUSE.

The distributions of nADC values measured using ss-EPI and MUSE are shown in Figure 4.6B. The mean nADC of malignant lesions was 0.69 ± 0.24 using ss-EPI and 0.72 ± 0.18 using MUSE. The mean nADC of benign lesions was 0.79 ± 0.21 using ss-EPI and 0.97 ± 0.23 using MUSE. There was no significant difference in nADC values of malignant and benign lesions measured using ss-EPI or MUSE ($p = 0.62$ and $p = 0.28$, respectively). The separation of the mean nADC values for malignant and benign lesions was greater using MUSE (0.25) than using ss-EPI (0.10).

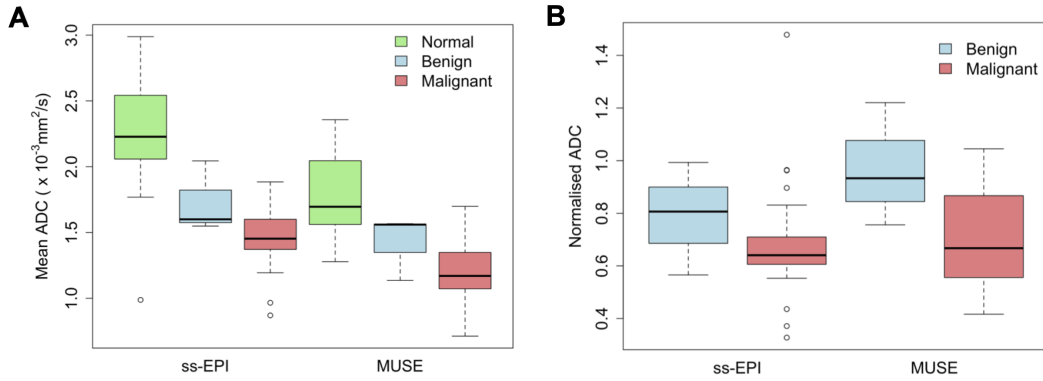


Figure. 4.6 Comparison of the A) ADC and B) nADC values measured using ss-EPI versus MUSE for malignant and benign lesions and normal fibroglandular tissue.

4.4 Discussion

In this study we demonstrated through phantom and clinical experiments that the quality of DWI can be improved using MUSE by significantly reducing blurring and distortion. ADC values measured in the phantom and in patients were in agreement with those in the literature [261].

Geometric distortion is prevalent in DWI due to the slow traversal through k -space along the phase-encoding direction when using ss-EPI. Accelerating the k -space trajectory through the use of a segmented-EPI technique such as MUSE can improve distortion compared to ss-EPI, as has been shown in previous studies implementing MUSE in the brain [262, 263]. Distortion has been evaluated in breast DWI by comparing lesion lengths in the anterior-posterior and left-right directions [255] or by using anatomic landmarks [257]. However, a more robust technique using the MI metric, a similarity measure often used in the field of image registration, was used by Teruel et al. when investigating the reduction of distortion after correction of B_0 inhomogeneities [264]. Distortion as measured by the MI metric in our study significantly improved using MUSE. In phantom studies, the best reduction in distortion was achieved using MUSE with 3 shots, however this was not clinically feasible due to the increase in acquisition time. A limitation of the MUSE technique is that the scan time increases proportional to the number of shots acquired. MUSE-DWI using simultaneous multi-slice excitation could be employed to reduce the acquisition time. This technique has been investigated in readout-segmented DWI in the breast [265].

Blurring is also a limiting factor in the clinical use of ss-EPI DWI, affecting accurate measurement of the ADC. Low resolution images result in ROIs that are averaged with adjacent normal fibroglandular tissue, leading to higher measured ADC values. While ADC values measured in the breast phantom did not vary substantially using ss-EPI or MUSE when using the same set of ROIs, ADC values measured in patients were lower using MUSE, suggesting that the difference in ADC may be a result of improved resolution and therefore improved ROI delineation using MUSE. More accurate quantification of the ADC should improve the separation of ADC values between malignant, benign and normal fibroglandular tissue, as is seen by the greater separation of nADC values between malignant and benign breast lesions. Wisner et al. similarly reported an improved separation of malignant and benign lesions using RESOLVE [254]. In our study, there were not enough lesions to do an analysis of diagnostic performance. However, given that readout-segmented techniques have been shown to have a better diagnostic performance compared to ss-EPI [58, 256], we expect that MUSE will achieve a better diagnostic performance than ss-EPI.

Less blurring and higher spatial resolution will allow DWI to better detect smaller lesions and satellite lesions. Adhering to the EUSOBI guidelines [266], the patients included in this study were referred for MRI for preoperative local staging of ipsilateral and contralateral newly diagnosed breast cancers where patients had dense breasts or lobular cancers (resulting in a high number of lobular cancers in this study which does not reflect the natural prevalence of this type of cancer) as well as for problem-solving (equivocal findings at mammography and ultrasound). In these cases, the improved image quality of MUSE-DWI can be used to better assess multi-focality and multi-centricity, which impacts surgical planning [96]. Reduced distortion may also allow for DWI to be used for morphological analysis, increasing the clinical utility of DWI.

There are limitations to this study. Firstly, low numbers of patients and lesions limit the conclusions that can be drawn. More benign lesions are required to evaluate diagnostic performance. Secondly, mostly mass type tumours were included in this study. Further work with a larger range of histopathological subtypes are required to prove the robustness of MUSE-DWI. Third, we were not able to acquire our current clinical protocol with a matrix size of greater than 96 x 128 to match the matrix size used for MUSE as this resulted in a high level of distortion. This limits the conclusions that can be drawn from the improvement in blurring. However, phantom images acquired with ss-EPI and MUSE at the same resolution allowed for more direct comparisons of blurring and distortion. Fourth, the readers were not blinded to the scan type and the images were read consecutively as opposed to allowing time between readings to forget the images. Finally, SNR and contrast-to-noise ratio (CNR) were

not compared due to the varying parallel acceleration factors and SENSE reconstructions used amongst acquisitions, resulting in different noise characteristics.

In conclusion, the image quality of MUSE-DWI was superior to that of ss-EPI-DWI and geometric distortion and blurring were significantly reduced, though there was an increase in acquisition time. MUSE-DWI is a promising technique for the detection and characterisation of breast lesions, however further work is required to evaluate the diagnostic performance of this technique using a larger range of histopathological subtypes of breast cancer.

Chapter 5

A meta-analysis comparing the diagnostic performance of abbreviated MRI and a full diagnostic protocol in breast cancer

The use of abbreviated MRI protocols (ABB-MRI) for the detection of breast cancer has gained increasing attention as these acquire a shortened version of a full diagnostic protocol (FDP-MRI) in a fraction of the time with comparable or reduced reading times, reducing the cost of the examination. However, before their adoption into mainstream clinical practice, equivalent diagnostic accuracy to a full diagnostic protocol must be demonstrated.

This chapter presents a comparative meta-analysis of the diagnostic performance of abbreviated and full diagnostic protocol MRI in the detection of breast cancer. Studies reporting abbreviated and full diagnostic protocols are compared in screening and enriched cohorts separately, and inconsistencies in study methodology are discussed.

Contents of this chapter have been published in *Clinical Radiology* [267] and presented at the International Society for Magnetic Resonance in Medicine (ISMRM) 2020 (abstract #2377).

5.1 Introduction

The use of abbreviated magnetic resonance imaging (ABB-MRI) protocols in the detection of breast cancer has gained increasing attention as these have substantially reduced image acquisition and interpretation times. The first prospective reader study of screening patients

using an abbreviated breast MRI protocol was reported by Kuhl et al. and showed equivalent diagnostic performance of ABB-MRI to a full diagnostic protocol (FDP) [70]. Initial studies created a short protocol from an existing dataset of standard breast MRIs and reported a reading study, generally showing equivalent performance to the standard acquisition [71, 268, 269]. More recent studies have created enriched cohorts of patients, for whom MRI has been used for problem solving or pre-operative staging, to assess the diagnostic accuracy of ABB-MRI in a robust manner with a sufficient number of cancers [270, 271]. Varying versions of the shortened protocols have been reported in these studies, with the general definition of an 'abbreviated protocol' using a non-contrast T₁-weighted (T₁W) sequence with at least one post contrast T₁W examination. Before the adoption of abbreviated MRI into mainstream practice it is important to ensure the shortened sequences gives equivalent diagnostic performance.

From a radiologist's perspective it is important to assess the use of ABB-MRI in a screening context and in a problem solving or pre-operative staging context separately. The advantage of an abbreviated protocol for screening is the ability to reduce healthcare costs, the time patients spend in the MRI scanner, as well as a reduced reading times for the radiologist. For problem solving and pre-operative staging, MRI is used extensively, however the case for abbreviated MRI for this clinical question is less compelling, as a full protocol is more likely to be more diagnostically useful. In order to adopt abbreviated MRI for screening, prospective trials need to be undertaken with careful comparison between abbreviated MRI and standard MRI protocols. However, in order to do this safely, assimilation of the evidence is required to show equivalence or at least non-inferiority using published data before a randomised trial is undertaken.

While several review articles have examined the protocols and diagnostic performances of published ABB-MRI studies [272–278], to date no meta-analysis has been performed that systematically compares the diagnostic performance of ABB-MRI with full diagnostic protocol MRI (FDP-MRI). This meta-analysis examines the evidence from screening only cohorts and separately from enriched cohorts.

5.2 Methods

This meta-analysis was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis for Diagnostic Test Accuracy (PRISMA-DTA) guidelines [279].

5.2.1 Literature search

PubMed and EMBASE databases were searched in August 2019 for studies assessing the diagnostic performance of abbreviated MRI protocols in the detection of breast cancer in either a screening or an enriched cohort of women. The patient population of screening studies consisted of screening mild-moderate or high-risk women, including women with a personal history of breast cancer. The patient population of enriched cohort studies included either combinations of screening, suspicious and known cancer cases or cases selected by the authors. The search strategy used was ((breast)) AND abbreviated) AND (MR OR MRI OR magnetic resonance imaging)). A full manual search of reference lists from all included studies was also undertaken.

5.2.2 Study selection

Studies were included if they met the following eligibility criteria: (1) published in a peer reviewed journal (abstracts and conference proceedings excluded), (2) in English, (3) the patient population was reported and included either a screening cohort or an enriched cohort of patients, (4) details of the full and abbreviated protocols were reported, (5) the diagnostic performance of both ABB-MRI and FDP-MRI in the detection of breast cancer was reported. Studies focusing on the development of an abbreviated protocol or technique were excluded.

5.2.3 Data extraction

The following information was obtained from studies: first author, publication year, prospective or retrospective study design, number of patients, number of cancers, ABB-MRI and FDP-MRI protocol sequences, number of readers and experience in years, examination times and reading times of ABB-MRI and FDP-MRI, and interval of time between reading ABB-MRI and FDP-MRI. The sensitivity and specificity of ABB-MRI and FDP-MRI protocols for each study was recorded. The number of true-positive (TP), false-negative (FN), false-positive (FP) and true-negative (TN) findings using ABB-MRI and FDP-MRI were either extracted from studies where reported or calculated from the number of included cancers. For studies that reported multiple readers, the number of TP/FN/FP/TN were extracted from only the first reader to ensure integer numbers of lesions for the meta-analysis. For studies that reported multiple ABB-MRI protocols, the diagnostic performance of the protocol that used a contrast-enhanced sequence and the smallest number of additional sequences was extracted.

5.2.4 Data quality assessment

The Quality Assessment of Diagnostic Accuracy Studies-2 was used to assess the risk of bias and concerns regarding applicability to the review question [179]. Risk of bias was assessed in four domains: patient selection (e.g. mild-, moderate- or high-risk patients for screening studies), appropriate index test (interpretation of ABB-MRI and FDP-MRI protocols without knowledge of final diagnosis, appropriate length of time or blinding between reading of ABB-MRI and FDP-MRI protocols), reference standard (use of histological analysis or follow-up), and flow and timing. The degree of heterogeneity between studies was assessed using Cochran's Q-test and the Higgins I^2 test [181]. A p-value of < 0.05 for Cochran's Q-test or an I^2 value of greater than 50% indicated statistically significant heterogeneity.

5.2.5 Statistical analysis

Forest plots of sensitivity and specificity for included studies were constructed. The bivariate model of Reitsma et al. [182] was used to estimate pooled sensitivities, specificities and areas under the curve (AUCs) for ABB-MRI and FDP-MRI on a per-lesion basis, and summary receiver operating characteristic (sROC) curves were constructed. Screening studies and enriched cohort studies were pooled separately to avoid bias. Additionally, the exam times, reading times, for all studies were compared using a paired t-test, with a p-value < 0.05 indicating a statistically significant result. The reported sensitivities and specificities of ABB-MRI and FDP-MRI for studies in the screening and enriched cohort groups were compared separately using a paired t-test for each comparison with a p-value < 0.05 indicating a statistically significant result. Analysis was performed using statistical software (R version 3.1.3; R Foundation for Statistical Computing, Vienna, Austria) using the *mada* package.

5.3 Results

5.3.1 Study selection and data extraction

The literature search of the PubMed and EMBASE databases returned 63 studies after removing duplicates. 30 studies were excluded after a review of the titles and abstracts. The full text of the remaining 33 studies was reviewed and 20 studies excluded as they did not meet the eligibility criteria. 13 studies (5 screening studies and 8 enriched cohort studies) were included in the meta-analysis [70–73, 81, 270, 271, 280–285]. One study was excluded as the patient population contained a subset of patients previously reported by the authors in a study included in the meta-analysis [286]. The study selection process is shown in Figure 5.1.

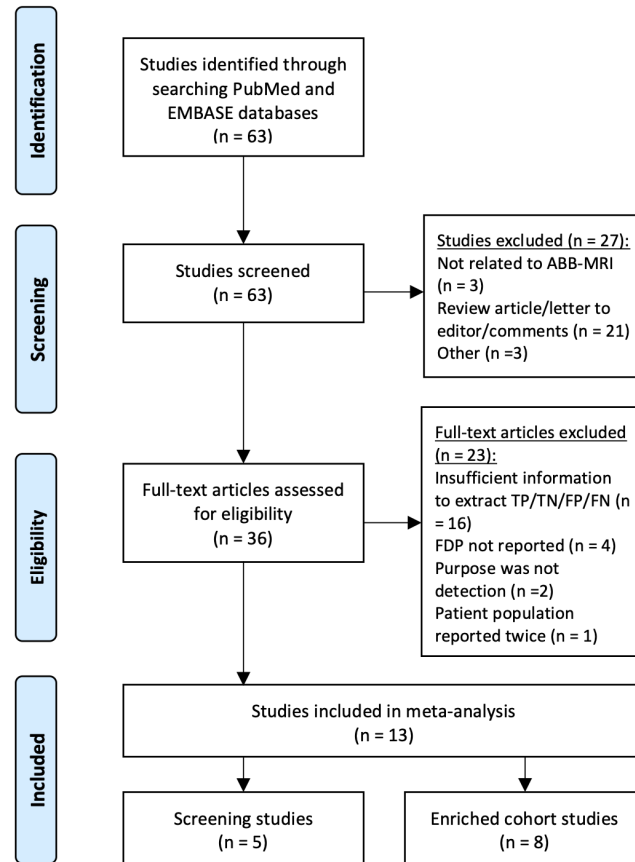


Figure. 5.1 PRISMA flow diagram for study selection and exclusion.

Details of included screening and enriched cohort studies are given in Tables 5.1 and 5.2, respectively. Screening studies included 2588 patients with 97 cancers. Enriched cohort studies included 1432 patients with 540 cancers. Technical details of included studies are given in Table 3. There was a large variation in patient population, study methodology and ABB-MRI protocols reported in included studies. All studies used at least one pre-contrast and one post-contrast sequence in their abbreviated protocol. The mean exam time was 7.4 minutes for ABB-MRI and 19.2 minutes for FDP-MRI ($p = 0.002$). The mean reading time was 1.4 minutes for ABB-MRI and 3.8 minutes for FDP-MRI ($p = 0.01$). The time between reading protocols ranged from immediately after to one month after. The majority of readers involved in studies had over 6 years of experience.

Figure 5.2 shows the results of QUADAS-2 assessment. For patient selection, some enriched cohort studies were found to have applicability concerns due to a combination of screening and patients with known cancers. For index tests, risks of bias found were due to either lack of reporting of the time between the reading of the ABB-MRI and FDP-MRI protocols (unclear risk) or the reading of the FDP-MRI directly after the ABB-MRI protocol (high risk). The use of a reference standard was unclear for one study. Regarding flow and timing, all studies were considered to have a low risk of bias.

Figure. 5.2 Results of quality assessment using QUADAS-2.

Table 5.1 Characteristics of included screening studies

Study (Year)	Study Design	Patient Population	No. of cancers (patients)	ABB-MRI						FDP-MRI					
				SEN (%)	SPE (%)	TP	FN	FP	TN	SEN (%)	SPE (%)	TP	FN	FP	TN
Dialani (2019)[72]	R	High risk	7 (259)	100	89	7	0	29	223	100	86	7	0	35	217
Panigrahi (2017)[283]	P	High risk	49 (1052)	82	97	40	9	27	976	82	97	40	9	26	977
Chen (2017)[280]	R	Mild risk/ dense breast	14 (356)	93	87	13	1	44	298	100	97	14	0	11	331
Chen (2017)[73]	R	Mild risk/ dense breast	16 (478)	94	88	15	1	54	408	100	95	16	0	25	437
Kuhl (2014)[70]	P	Mild risk/ dense breast	11 (443)	100	94	11	0	24	408	100	95	11	0	22	410

R: Retrospective, P: Prospective, ABB-MRI: abbreviated MRI, FDP-MRI: full diagnostic protocol MRI, SEN: sensitivity, SPE: specificity, TP: true positive, FN: false negative, FP: false positive, TN: true negative

Table 5.2 Characteristics of included enriched cohort studies

Study (Year)	Study Design	Patient Population	No. of cancers M/B (patients)	ABB-MRI						FDP-MRI					
				SEN (%)	SPE (%)	TP	FN	FP	TN	SEN (%)	SPE (%)	TP	FN	FP	TN
Seppala (2018) [284]	R	Selected cases	23/77 (100)	65	79	15	8	16	61	83	79	19	4	16	61
Oldrini (2017)[271]	R	Pre-operative staging, high risk screening, problem solving, nipple discharge	58/48 (70)	93	60	54	4	19	29	93	60	54	4	19	29
Romeo (2017)[285]	R	Known cancer, suspicious mammogram /ultrasound, pre-operative staging	110/70 (98)	99	93	109	1	5	65	97	95	107	3	3	67
Petrillo (2017)[281]	R	Consecutive cases	207/301 (508)	99	68	206	1	96	205	99	69	206	1	93	208
Machida (2017)[282]	R	Selected cases	31/60 (88)	87	92	27	4	5	55	87	90	27	4	6	54
Moschetta (2017)[270]	R	High risk screening, suspicious mammogram, pre-operative staging	75/110 (470)	89	91	67	8	10	100	92	92	69	6	9	101
Bickelhaupt (2016)[81]	P	Suspicious mammogram	24/26 (50)	88	92	21	3	2	24	92	92	22	2	2	24
Grimm (2015)[71]	R	Selected cases	12/12 (48)	92	69	11	1	4	8	92	58	11	1	5	7

R: Retrospective, P: Prospective, M: malignant, B: benign, ABB-MRI: abbreviated MRI, FDP-MRI: full diagnostic protocol MRI, SEN: sensitivity, SPE: specificity, TP: true positive, FN: false negative, FP: false positive, TN: true negative

Table 5.3 Technical details of included studies

Study (Year)	Vendor Field Strength	ABB-MRI protocol	FDP-MRI protocol	Examination Time (minutes)		Number of readers (experience)	Reading Time (minutes)		Time between reading protocols
				ABB-MRI	FDP-MRI		ABB-MRI	FDP-MRI	
Dialani (2019)[72]	GE, Siemens 1.5T	T ₁ W pre-contrast and first post-contrast, MIP, T ₂ W	T ₁ W pre-contrast, 4 phases post-contrast, MIP, T ₂ W	8	16	5 (4-28 years)	2.4		Different readers used
Seppala (2018)[284]	GE 1.5T	FAST-MIP (Pre-contrast and first post-contrast, subtraction, MIP)	T ₂ W, T ₁ W, pre-contrast and 4 post-contrast T ₁ W, 3D FAME, subtraction, MIP	4.6	16.75	3 (6-22 years)	1.2	2.3	Immediately after
Panigrahi (2017)[283]	GE 1.5T	Fat-saturated T ₁ W, first post-contrast subtraction, MIP	T ₂ W, T ₁ W, pre-contrast and 3 post-contrast T ₁ W, delayed post-contrast T ₁ W, subtractions, MIP	3	24	10 (1-22 years)	2.4	6.6	Immediately after
Chen (2017)[280]	Siemens 3T	FAST-MIP (Pre-contrast and first post-contrast, subtraction, MIP)	T ₁ W, transverse and sagittal T ₂ W, DWI, pre-contrast and 5 post-contrast T ₁ W, subtraction, MIP, ADC map	10.5	32	2 (10 years)	0.6	3.5	1 month
Oldrini (2017)[271]	GE 1.5T	FAST (first post-contrast, subtraction), T ₂ W	T ₁ W, T ₂ W, pre-contrast and 4 post-contrast VIBRANT T ₁ W, TRICKS during contrast injection, subtraction, MIP	5.8	15	2 (5 and 6 years)	1.5	5.4	2 weeks
Romeo (2017)[285]	Philips 1.5T	Subtraction images at 28s, 84s and 252s after injection	T ₂ W, pre-contrast and 5 post-contrast T ₁ W, subtraction, T ₁ W isovolumetric	7	15.8	2 (8 and 16 years)	1	1	20-30 days
Petrillo (2017)[281]	Aurora 1.5T	Pre-contrast and first post-contrast subtraction, MIP	T ₂ W, pre-contrast and 4 post-contrast T ₁ W, subtraction, MIP			2 (20 years)			
Machida (2017)[282]	Siemens 3T	Pre-contrast and 80s post contrast VIBE with TWIST	3D VIBE, pre-contrast and 4 post-contrast VIBE with TWIST, T ₁ W, T ₂ W, DWI, ADC map, subtraction, MIP			2 (14 and 15 years)			>1 week
Moschetta (2017)[270]	Philips 1.5T	STIR, T ₂ W, pre-contrast and post-contrast THRIVE (3 mins after injection)	STIR, T ₂ W, pre-contrast and 5 post-contrast THRIVE	10	16		2	6	1 week
Chen (2017)[73]	Siemens 3T	FAST-MIP (Pre-contrast and first post-contrast, subtraction, MIP)	Axial T ₁ W, T ₂ W, sagittal T ₂ W, pre-contrast and 7 post-contrast T ₁ W, subtraction, MIP	13		2 (10 years)	0.75	3	1 month
Bickelhaupt (2016)[81]	Philips 1.5T	T ₂ W, T ₁ W first post-contrast subtraction, MIP	T ₁ W, T ₂ W, T ₂ W with SPAIR, DWIBS, pre-contrast and 5 post-contrast T ₁ W, T ₁ W, subtraction, MIP	<5		2 (2 and 6 years)	0.5		Different readers used
Grimm (2015)[71]	Siemens 1.5T	T ₂ W, T ₁ W first post-contrast subtraction	T ₁ W, T ₂ W, pre-contrast and 3 or 4 post-contrast T ₁ W, subtraction	11	20	2 (7 and 14 years)	2.98	2.95	1 month
Kuhl (2014)[70]	Philips 1.5T	FAST-MIP (Pre-contrast and first post-contrast, subtraction, MIP)	Pre-contrast and 5 post-contrast T ₁ W, T ₂ W, T ₁ W, MIP	3	17	2 (6 and 18 years)	0.5		Immediately after

ABB-MRI: abbreviated MRI, FDP-MRI: full diagnostic protocol MRI, T₁W: T₁-weighted, T₂W: T₂-weighted, MIP: maximum intensity projection, FAME: enhanced fast gradient echo 3D, FAST: first post-contrast subtracted, DWI: diffusion-weighted imaging, ADC: apparent diffusion coefficient, VIBRANT: volume image breast assessment, TRICKS: time-resolved imaging of contrast kinetics, VIBE: volumetric interpolated breath-hold examination, TWIST: time-resolved angiography with stochastic

5.3.3 Statistical analysis

The results of pooled analysis are given in Table 5.4. Low heterogeneity was measured between studies using enriched cohorts. I^2 values of 0% were measured for screening studies using both ABB-MRI and FDP-MRI, however this was due to an insufficient number of studies included to use this technique as opposed to lack of heterogeneity.

Forest plots for sensitivity and specificity are shown in Figure 5.3. For screening studies, the confidence intervals are large, and are much larger for sensitivity than for specificity due to the very small number of cancers in proportion to normal cases ($n = 97$ for 2588 patients for all screening studies combined). For enriched cohort studies with a more balanced number of cancers and normal cases, the confidence intervals are more similar, though the confidence intervals are still large overall.

Summary receiver operating characteristic curves are shown in Figure 5.4. FDP-MRI achieved a higher sensitivity, specificity and AUC than ABB-MRI for both screening and enriched cohort studies. The difference in diagnostic performance between ABB-MRI and FDP-MRI was lower for enriched cohort studies. However, the sensitivities and specificities of ABB-MRI and FDP-MRI were not statistically significantly different for screening studies or enriched cohort studies ($p = 0.18$ and 0.27 , $p = 0.18$ and 0.93 , respectively). The pooled AUC for ABB-MRI was the same for screening and enriched cohorts.

Table 5.4 Results of pooled analysis

Type of study	No. of studies	No. of patients	No. of cancers	Protocol	Heterogeneity		Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
					Cochran's Q p-value	I ² (%)			
Screening	5	2588	62	ABB-MRI	0.80	0	90 (79 – 96)	92 (86 – 95)	0.94 (0.90 – 0.98)
				FDP-MRI	0.56	0	92 (77 – 97)	95 (91 – 97)	0.97 (0.97 – 0.99)
Enriched	8	1432	540	ABB-MRI	0.31	15	93 (83 – 97)	83 (73 – 90)	0.94 (0.87 – 1.00)
				FDP- MRI	0.34	12	93 (88 – 96)	84 (72 – 91)	0.95 (0.92 – 1.00)

ABB-MRI: abbreviated MRI, FDP-MRI: full diagnostic protocol MRI, CI: confidence interval, AUC: area under the receiver operating characteristic curve

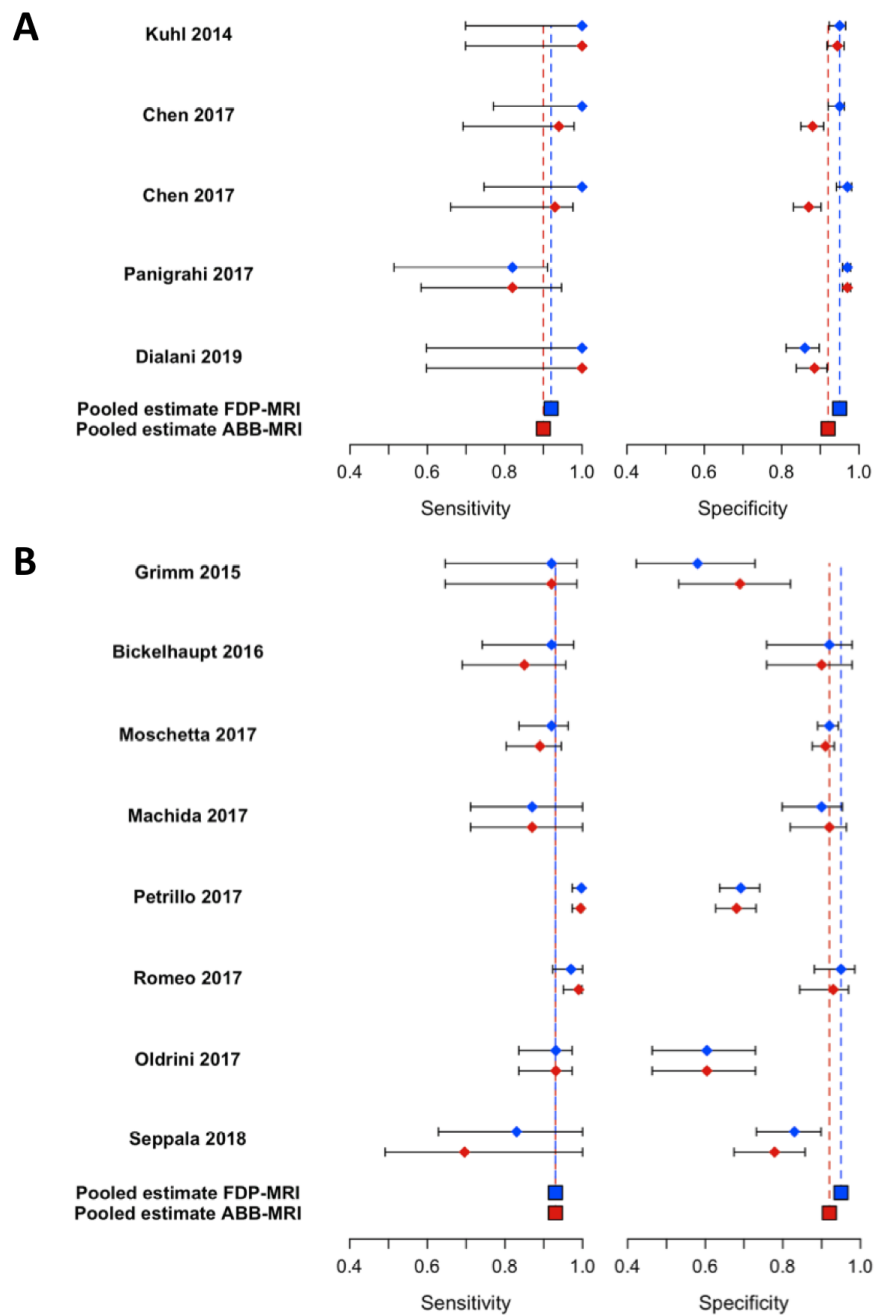


Figure. 5.3 Forest plots of the sensitivity and specificity of full diagnostic protocol MRI (FDP-MRI) and abbreviated MRI (ABB-MRI) for included A) screening and B) enriched cohort studies with 95% confidence intervals. Vertical lines denote pooled summary estimates of sensitivity and specificity.

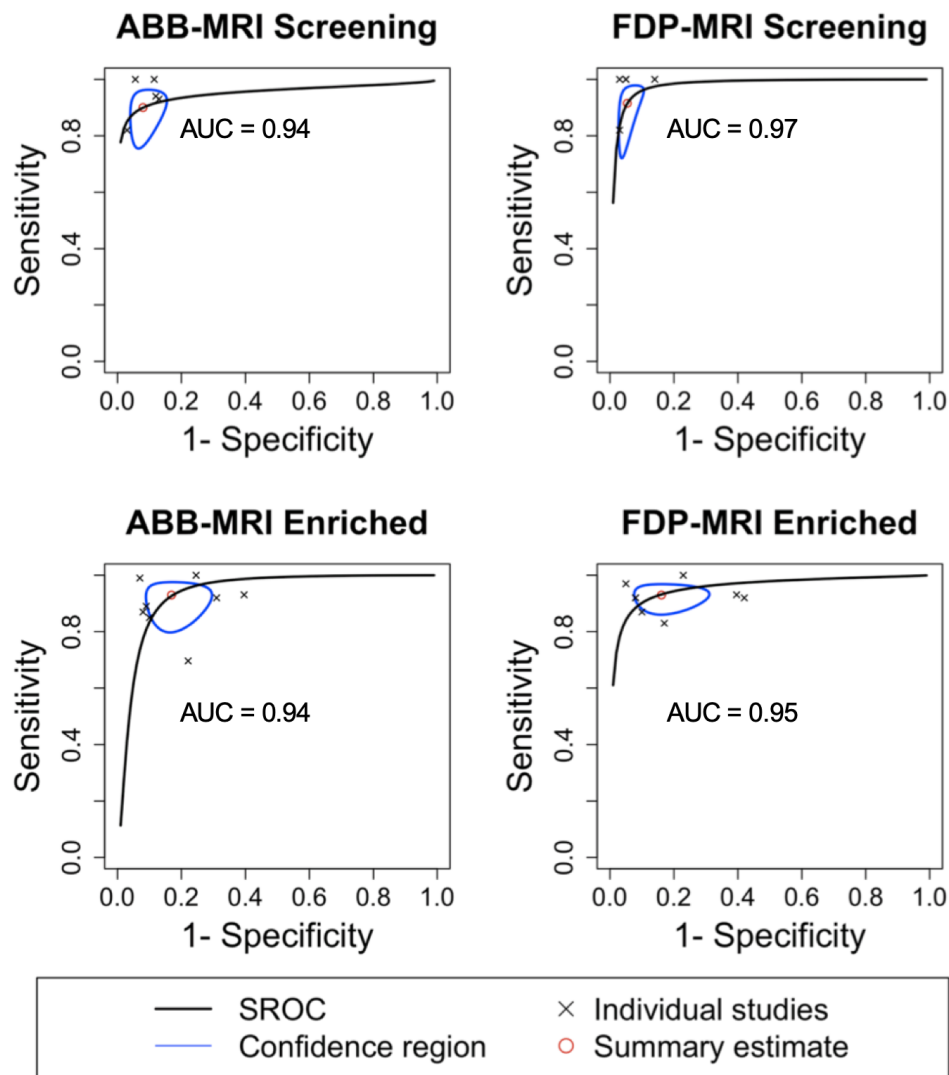


Figure. 5.4 Summary receiver operating characteristics (ROC) curves for abbreviated MRI (ABB-MRI) and full diagnostic protocol MRI (FDP-MRI) protocols using the bivariate model with 95% confidence regions. The pooled AUCs of ABB-MRI and FDP-MRI for screening studies were 0.94 and 0.97, respectively. The pooled AUCs of ABB-MRI and FDP-MRI for enriched cohort studies were 0.94 and 0.95, respectively.

5.4 Discussion

This meta-analysis showed that in a screening setting, the diagnostic accuracy of abbreviated MRI was lower but not statistically significantly different to the full diagnostic protocol MRI (pooled AUCs 0.94 and 0.97, respectively). For studies that used enriched cohorts, the

performance of abbreviated MRI was similar to that of the standard protocol (pooled AUCs 0.94 and 0.95, respectively).

Comparison and pooling of ABB-MRI studies through a meta-analysis is complicated by the variation in patient populations reported as sensitivity and specificity performance can be altered by the expected prevalence of cancers in the cohort. It is better not to group abbreviated protocols used for screening and for other clinical indications together. Amongst screening studies, Kuhl et al. and Chen et al. reported results from screening mild- or moderate-risk patients [70, 73], whereas Panigrahi et al. and Dialani et al. reported results from screening high-risk patients [283, 72]. The effective rate of detected cancers will differ between these two groups, and therefore it may not be meaningful to pool their diagnostic performances. Furthermore, though there was a variation in patient populations, an I^2 of 0% was measured between screening studies indicating no heterogeneity. However, it has been shown that conclusions of low heterogeneity for a meta-analysis with a small number of studies are unjustified as confidence intervals for these heterogeneity estimates are large [181, 287]. Results from screening studies may also be underpowered due to the large number of normal cases, where specificity will be inherently comparable for ABB-MRI and FDP-MRI and the low number of cancer cases results in sensitivity values with large confidence intervals. Amongst enriched cohort studies, Moschetta et al. reported a cohort of combined screening, problem solving and preoperative staging patients [270]. Bickelhaupt et al. reported a cohort of patients with suspicious mammograms [81]. Grimm et al. reported a cohort of selected cases with a balanced number of cancers and benign and normal cases [71], though the readers were blinded to the percentage of each case. It is unclear what effect these combinations of patients within a population would have on reading images. While enriched cohorts were able to demonstrate equivalent performance to a full diagnostic protocol, they do not reflect the clinical setting of interest and may not be applicable in a screening setting.

Other than differences in patient population, the assessment of the quality of studies included in the meta-analysis using QUADAS-2 highlighted other variations in study design. Given the claims of equivalent diagnostic performance to standard protocols, it is important to scrutinise the methodology of these reader studies before it is possible to safely adopt abbreviated MRI into clinical practice. While some studies left up to a month between reading images from different protocols, some read the full protocol directly after the abbreviated protocol. This may be appropriate when assessing changes in management with the addition of extra sequences, however both protocols must be tested equally to robustly compare the diagnostic performance of ABB-MRI and FDP-MRI. Given that most studies were performed by readers with many years of experience, it may be that the high diagnostic accuracy and faster reading times achieved using abbreviated protocols would not be possible with less

experienced readers. Furthermore, readers in retrospective studies would not be afraid of misdiagnoses and may perform differently when reading images in a real clinical setting. Only three of the studies included were prospective studies, and larger prospective and multi-centre trials with defined inclusion criteria are required to validate the performance of ABB-MRI in a purely screening setting. The lack of precision in pooled estimates also necessitates large prospective trials, given that the lower end of the ranges of the sensitivity and specificity of ABB-MRI in a screening setting (79% and 86%, respectively) are not good enough to be used in a screening situation and unlikely to be cost-effective. There are multiple ongoing prospective studies, the largest being the multi-centre EA1141 trial (Comparison of AB-MRI and DBT in Breast Cancer Screening in Women with Dense Breasts), finding a higher rate of invasive cancer detection using ABB-MRI compared to digital breast tomosynthesis (DBT) in a screening cohort of 1444 women with dense breasts and only mild to moderate risk of breast cancer [288, 77].

The various reported ABB-MRI protocols have been previously reviewed [272, 275, 277]. In this meta-analysis, only one set of reported sensitivity and specificity values were extracted from each study to avoid overrepresentation of a sample, although many studies have compared the diagnostic performance of multiple combinations of sequences to investigate the added value of extra sequences in increasing specificity and confidence in diagnosis. Overall, studies have dropped the full dynamic time course in order to save time, opting for one pre-contrast and one post-contrast time point. Grimm et al. found that the addition of a second post-contrast time point did not improve diagnostic accuracy [71]. Different studies added either a T₂-weighted (T₂W) sequence or a diffusion-weighted examination to complement the contrast examination. Dialani et al. found that the addition of a T₂W sequence did not result in a significant change in management [72]. A second abbreviated protocol including a diffusion-weighted imaging (DWI) sequence was used by Bickelhaupt et al. which performed better than the protocol using only contrast-enhanced images, matching the accuracy of the full diagnostic protocol [81]. Chen et al. also found that the addition of DWI improved sensitivity and specificity [73]. There is a growing interest in non-contrast-enhanced screening and DWI is increasingly used in the detection of breast cancer, with advanced DWI techniques showing a high sensitivity and specificity in the characterisation of breast lesions [67, 50].

This study has several limitations. Firstly, there were a low number of studies contributing to the pooled estimates resulting in relatively wide confidence intervals, particularly for screening studies. Second, there were many studies that investigated the diagnostic performance of ABB-MRI but did not perform a reader study for the full diagnostic protocol and were therefore not included in the meta-analysis. It has been shown that different results are obtained when pooling non-comparative studies (evaluating only one test) and comparative

studies (evaluating both tests equally) [289]. As such, robustly designed comparative studies where all patients received both tests under the same conditions were preferred. Third, while separate pooled analysis was carried out for screening and enriched cohort studies, there were still variations in patient populations within these groups. Fourth, it was unclear if there was an overlap between patient populations in two studies (both by Chen et al. [280, 73]) which could result in overrepresentation of a sample in pooled estimates, though the full protocols reported were sufficiently different. The authors could not be contacted for clarification. Fifth, as each study population could be used only once, the meta-analysis did not incorporate the potential added value of additional sequences that were investigated in some studies.

In conclusion, this meta-analysis of 13 studies found that abbreviated MRI had an overall high diagnostic performance in the detection of breast cancer. The diagnostic performance was similar to that of a full diagnostic protocol amongst enriched cohorts and was lower but not significantly different in a screening setting. While acquisition and interpretation times were significantly reduced compared to a full diagnostic protocol, there was a variation in study methodology and sequences chosen, limiting the conclusions that can be drawn. Further large prospective multicentre trials are required to validate ABB-MRI in a real screening environment.

Chapter 6

Predicting pathological complete response to neoadjuvant chemotherapy in breast cancer: machine learning-based analysis of radiomics features derived from pre-treatment DCE-MRI

In this chapter, machine learning models trained on radiomics features derived from pre-treatment dynamic contrast-enhanced MRI (DCE-MRI) are used to predict pathological complete response (pCR) to neoadjuvant chemotherapy. The stability of feature selection using logistic regression classification with the least absolute shrinkage and selection operator (LASSO) is assessed and the ability of features derived from individual time points in the dynamic series to predict pCR is investigated.

Contents of this chapter have been presented at the International Society for Magnetic Resonance in Medicine (ISMRM) 2020 (abstract #0573).

6.1 Introduction

Neoadjuvant chemotherapy (NACT) is used in the management of breast cancer to reduce the size of tumours before surgery, improving the rate of breast conservation surgery and reducing the extent of axillary surgery [290]. The best outcome from NACT is pathological complete response to chemotherapy (pCR), defined as the absence of residual invasive carcinoma in the breast or lymph nodes, which has been shown to be an indicator of long-term survival

and is associated with an 80% reduction in recurrence rate [291, 292]. MRI is the most accurate technique for the assessment of response to treatment [103, 293]. In particular, dynamic contrast-enhanced MRI (DCE-MRI) can be used to measure levels of contrast enhancement in different types of tissue that reflect changes in tumour vascularity, as well as providing high resolution anatomical information to assess morphology and accurately measure changes in tumour volume. Assessment of response during treatment is crucial to avoid unnecessary toxicity and cost by stopping treatment or changing to a more effective regime if a patient is not showing therapeutic response or if a patient has achieved pCR. Early prediction of response at the start of chemotherapy could provide a more personalised approach to treatment, allowing for better timing of surgery, choice of treatment regimen and a more individualised prognosis.

Radiomics involves the extraction of a large number of quantitative features from images used to train a machine learning model for classification and is increasingly used in the prediction of pCR in breast cancer. DCE-MRI is commonly used for radiomics analysis as the high spatial resolution is ideal for the extraction of texture features that quantify the distribution of voxel grey levels in a tumour volume, providing a measure of intratumoural heterogeneity which is an established biomarker of poor prognosis [138]. Combining the kinetic, textural, and morphological features derived from DCE-MRI, as well as clinicopathological features, using machine learning approaches can provide a more sophisticated method of predicting pCR. Studies investigating the prediction of pCR from pre-treatment MRI using radiomics analysis to date have achieved a moderate accuracy (AUCs 0.69-0.83) [141, 144–147]. As well as accurately predicting pCR, there is interest in finding a stable set of informative and predictive radiomics features that add to our understanding of tumour response and could inform decisions about types of treatment. However, the link between texture features and underlying tumour pathology is not well understood. Where traditional statistics approaches look for associations between imaging features and outcomes, a machine learning model trained on large quantities of data aims to predict outcomes without focusing on the predictors used, limiting the information that can be gained from interpreting the features used in the model.

This study aimed to investigate the use of different types of image-derived features extracted from pre-treatment DCE-MRI, using machine learning techniques, to predict pCR in breast cancer. In addition, the predictive performance of features derived from different time points through the DCE-image series were compared to investigate the effect of dynamic contrast enhancement on the selection of radiomics features.

6.2 Methods

Patient selection

The local institutional review boards and ethics committees approved this prospective study and written informed consent was obtained from all patients (National Research Ethics Committee number 13/LO/0411). Patients were eligible if they had pathologically confirmed invasive breast cancer and neoadjuvant chemotherapy was planned as the first line of treatment. Between February 2014 and March 2019, 172 females were recruited to this study. 4 patients withdrew from the study and 11 patients did not receive neoadjuvant chemotherapy and went directly to surgery. 3 patients were excluded as they did not receive a consistent MRI protocol and 2 patients were excluded due to having rare types of invasive carcinoma (invasive spindle cell carcinoma and a cystic invasive ductal carcinoma) that made it difficult to either outline their lesion or the lesion was not suited for radiomics feature extraction. After exclusions, a total of 152 patients were included in the analysis (mean age 51.2 ± 10.7 years, range 28-75 years).

MRI acquisition

All MR examinations were performed before the start of neoadjuvant chemotherapy using a 3T system (MR750, GE Healthcare, Wakesha, WI) with a dedicated eight-channel breast array coil. Dynamic contrast-enhanced (DCE) MRI was acquired using a three-dimensional fast spoiled gradient echo technique with *k*-space data sharing (volume image breast assessment-time-resolved imaging of contrast kinetics, or VIBRANT-TRICKS) [169, 170]. DCE-MRI was performed with TR/TE = 7.1/3.8ms, field-of-view = 350 mm, image matrix = 512×512 , section thickness = 1.4mm (interpolated), voxel size = $0.6 \times 0.6 \times 1.4\text{mm}^3$, flip angle = 12° , array spatial sensitivity encoding factor = 2.5 (phase direction), fat suppression = spatial-spectral water excitation, acquisition time = 8 minutes and 7 seconds.

Contrast agent was delivered using a catheter placed in the antecubital vein. A dose of 0.1mmol of gadopentetate dimeglumine (Magnevist, Bayer Schering, Berlin, Germany) per kilogram of body weight was administered using an automated injector at a rate of 3.0ml/s and was followed by a 25ml saline flush. 5 baseline volumes were acquired before contrast injection, followed by 43 contrast-enhanced volumes with a nominal temporal resolution of ~ 10 seconds. 5 subtraction volumes were generated by subtracting pre-contrast volumes from post-contrast volumes acquired approximately 1 min, 2.5 mins, 4 mins, 5.5 mins and 7 mins after contrast injection.

Histological analysis and response assessment

Tumour type, histological grade, hormone receptor status and HER2 expression were obtained from pathology reports of core biopsies performed before the start of chemotherapy. Histological grade was assessed using the Nottingham grading system [16]. As the progesterone receptor status was not assessed for a number of patients in this study ($n = 62$), patients were grouped into 4 immunohistochemical (IHC) molecular subtype groups given by Onitilo et al. based on hormonal receptor status (ER and PR) and HER2 expression: ER+PR+/HER2+, ER+PR+/HER2-, ER-PR-/HER2+, or ER-PR-/HER2- (i.e. triple negative) [294]. Histopathological assessment of response to chemotherapy was performed by expert pathologists after surgical excision following the last cycle of chemotherapy. Tumour response was classified using one of three categories given by Pinder et al. : 1) pathological complete response (pCR) either i) no residual carcinoma or ii) no residual tumour with the presence of DCIS, 2) partial response with either i) minimum residual disease ($<10\%$ of tumour remaining) or ii) $10\text{--}50\%$ residual tumour cellularity with evidence of response to therapy or iii) $>50\%$ residual tumour cellularity with some features of response to therapy, or 3) no evidence of response [295]. Lesions achieving partial response and no response were grouped as incomplete pathological response (non-pCR) for the purposes of this study. 60 patients achieved pCR and 92 patients did not achieve pCR (non-pCR).

Region of interest analysis

Regions of interest were manually delineated slice by slice on the non-subtracted second post-contrast time point volume by three breast radiologists (1, 4, and 5 years experience in breast imaging) using Osirix (version 8.0.1, Pixmeo, Switzerland) and propagated to other time points. One pathologically confirmed breast lesion was included per patient to avoid inter-lesion correlation that may confound results. For multi-focal and multi-centric lesions, the largest single index lesion was used. Volumes were registered to the second post-contrast time point using the simple ITK toolbox (<http://www.simpleitk.org>) [296, 297] if patient motion between time points was observed.

Radiomics feature extraction

The radiomics analysis workflow is shown in Figure 6.1. Before radiomics features were extracted, bias correction was performed for all MRI images to correct for low frequency intensity non-uniformity [298]. A total of 384 features were extracted for each lesion. For each of the 5 subtraction volumes, 74 features were calculated (18 first order histogram features, 24 Grey Level Co-occurrence Matrix (GLCM) features, 16 Grey Level Run Length

Matrix (GLRLM) features and 16 Grey Level Size Zone Matrix (GLSZM) features). 14 shape features were also extracted from each region of interest. Feature extraction was carried out using the PyRadiomics toolbox [299].

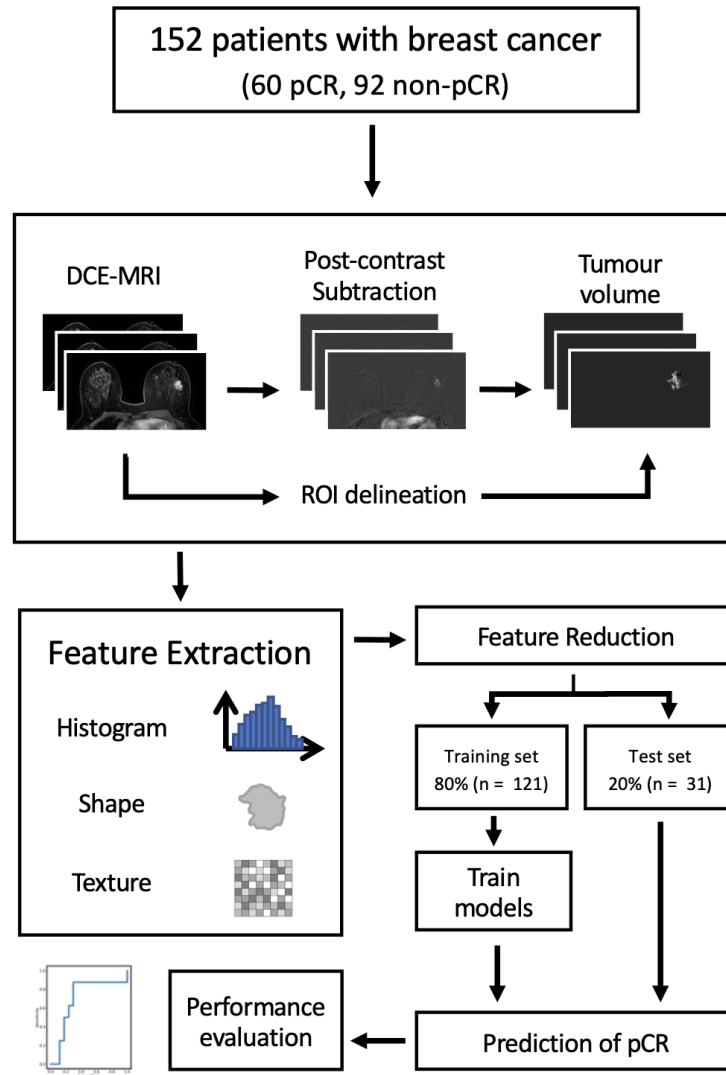


Figure. 6.1 Radiomics analysis workflow.

Machine learning predictive modelling

Data was split in a patient-wise fashion into 80%/20% independent training and test sets, maintaining the proportion of pCR and non-pCR cases in each set. Radiomics features were normalised using z-score standardisation before training. Logistic regression classification

Table 6.1 Features sets used to train logistic regression models

Feature set	Features Included	N features
Shape + Histogram (S+H)	14 Shape features 18 Histogram features per subtraction	104
Shape + Histogram + Texture (S+H+T)	14 Shape features 18 Histogram features per subtraction 56 Texture features per subtraction	384
Shape + Histogram + Pathology (S+H+P)	14 Shape features 18 Histogram features per subtraction 4 Clinicopathological features	108
Shape + Histogram + Texture + Pathology (S+H+T+P)	14 Shape features 18 Histogram features per subtraction 56 Texture features per subtraction 4 Clinicopathological features	388

models using the least absolute shrinkage and selection operator (LASSO) to enable feature selection [300] were trained with a 5-fold stratified cross-validation scheme. Grid search was used to tune the regularisation hyperparameter, C . The predictive performance of models was then evaluated on the test set using receiver operating characteristic curve (ROC) analysis. The model that achieved the highest cross-validation area under the curve (AUC) score in the training set in predicting pCR was reported.

Feature sets

To compare the ability of different types of radiomics features to predict pCR, models were trained using feature sets containing the shape (S), histogram (H), texture (T) and clinicopathological (P) features. The combinations of features in each feature set are summarised in Table 6.1. The 4 clinicopathological variables included age, grade, tumour subtype, and molecular IHC subtype group.

Feature reduction and selection

To reduce the dimensionality of the data before training, redundant correlated features were removed from feature sets as LASSO can be unstable when features are strongly correlated, and for several correlated variables, LASSO might select one of them randomly [300]. A correlation matrix was generated for all 388 radiomics features using Spearman's rank correlation coefficient (ρ). For each pair of features that were strongly correlated (a value of $|\rho| > 0.95$), ROC analysis was performed for each feature independently in the prediction of pCR and the redundant feature with the lower AUC was removed from the feature set. To

assess the ability of LASSO to handle collinearity, the features selected by models trained on full feature sets and reduced feature sets were compared.

Comparison of time points

To investigate the ability of each post-contrast time point to predict pCR, models were trained using image-derived features from each time point individually (88 features per time point) as well as the image-derived features plus the 4 clinicopathological features (92 features per time point). The performances of the models and the features selected were compared. To investigate the added value of each post-contrast time point, models were trained using features from the first two time points, the first three time points and the first four time points. The performances of the models were compared to the best performing models trained using features from only the first post-contrast time point and features from the full dynamic series.

Statistical analysis

The Mann-Whitney U-test was used to compare differences in all radiomics features between the pCR and non-pCR groups. Correction for multiple comparisons was performed using the Benjamini-Hochberg correction with a false discovery rate of 0.05 [301]. Clinicopathological variables were compared in pCR and non-pCR groups using the Mann-Whitney U-test for continuous variables and the χ^2 test for categorical variables. p-values < 0.05 were considered statistically significant. 95% confidence intervals for cross-validation scores and test scores were calculated using bootstrapping with 1000 samples. All analysis was carried out using python 3.7.0 (<https://www.python.org>) and the software modules numpy, pandas, scipy, matplotlib, and sklearn [302].

6.3 Results

Clinical characteristics

Patient and lesion characteristics for the training and test sets are given in Table 6.2.

Predictive modelling

A summary of the 18 models trained using different combinations of features and post-contrast time points is given in Figure 6.2. Figure 6.3A shows the change in cross-validation score for each feature set with regularisation parameter, C. For models trained using S+H features, the cross-validation score reached a maximum at 19 features then the performance of the models decreased with increasing number of features included in the model. For models trained using S+H+P, S+H+T, and S+H+T+P features, the cross-validation score reached a maximum then remained stable with increasing number of features included in the model.

Table 6.2 Patient and lesion characteristics

	Training set (n = 121)	Test set (n = 31)	p-value*
Mean age in years \pm SD (range)	51.2 \pm 10 (28-75)	51.2 \pm 12 (28 - 72)	0.49
	N (%)		
Pathological response			
pCR	48 (39.7)	12 (38.7)	0.91
non-pCR	73 (60.3%)	19 (61.3)	
Tumour histology			
Invasive ductal carcinoma	108 (89.3)	29 (93.6)	0.57
Invasive lobular carcinoma	3 (2.5)	1 (3.2)	
Invasive micropapillary carcinoma	2 (1.7)	1 (3.2)	
Invasive ductal carcinoma with micropapillary features	2 (1.7)		
Metaplastic carcinoma	1 (0.8)		
Mixed ductal and lobular carcinoma	1 (0.8)		
Pleomorphic invasive lobular carcinoma	1 (0.8)		
Mixed invasive ductal and mucinous carcinoma	1 (0.8)		
Invasive mucinous carcinoma	1 (0.8)		
Metaplastic carcinoma with squamous differentiation	1 (0.8)		
Histological grade			
2	29 (24.0)	6 (19.4)	0.76
3	92 (76.0)	25 (80.6)	
Molecular subtype group			
ER+PR+/HER2+	33 (27.3)	7 (22.6)	0.77
ER+PR+/HER2-	30 (24.8)	6 (19.3)	
ER-PR-/HER2+	14 (11.6)	5 (16.1)	
ER-PR-/HER2-	44 (36.3)	13 (42.0)	

SD: standard deviation, pCR: pathological complete response, ER: estrogen receptor, PR: progesterone receptor, HER2: human epidermal growth factor receptor 2, + : positive, - : negative * Clinicopathological variables were compared in training and test sets using the Mann-Whitney U-test for continuous variables and the χ^2 test for categorical variables.

The models trained using feature sets including clinicopathological features both had a peak in cross-validation score with 1 feature selected (molecular subtype group) with the same performance (cross-validation score 0.70), however a higher cross-validation score was achieved using the S+H+P feature set by including 18 additional features.

The cross-validation scores, test scores, and number of features selected by models trained on each set of features are shown in Table 6.3. The cross-validation score of models trained on S+H features was higher than those trained on S+H+T features. The highest cross-validation and test score was achieved using a model trained on S+H+P features. 19

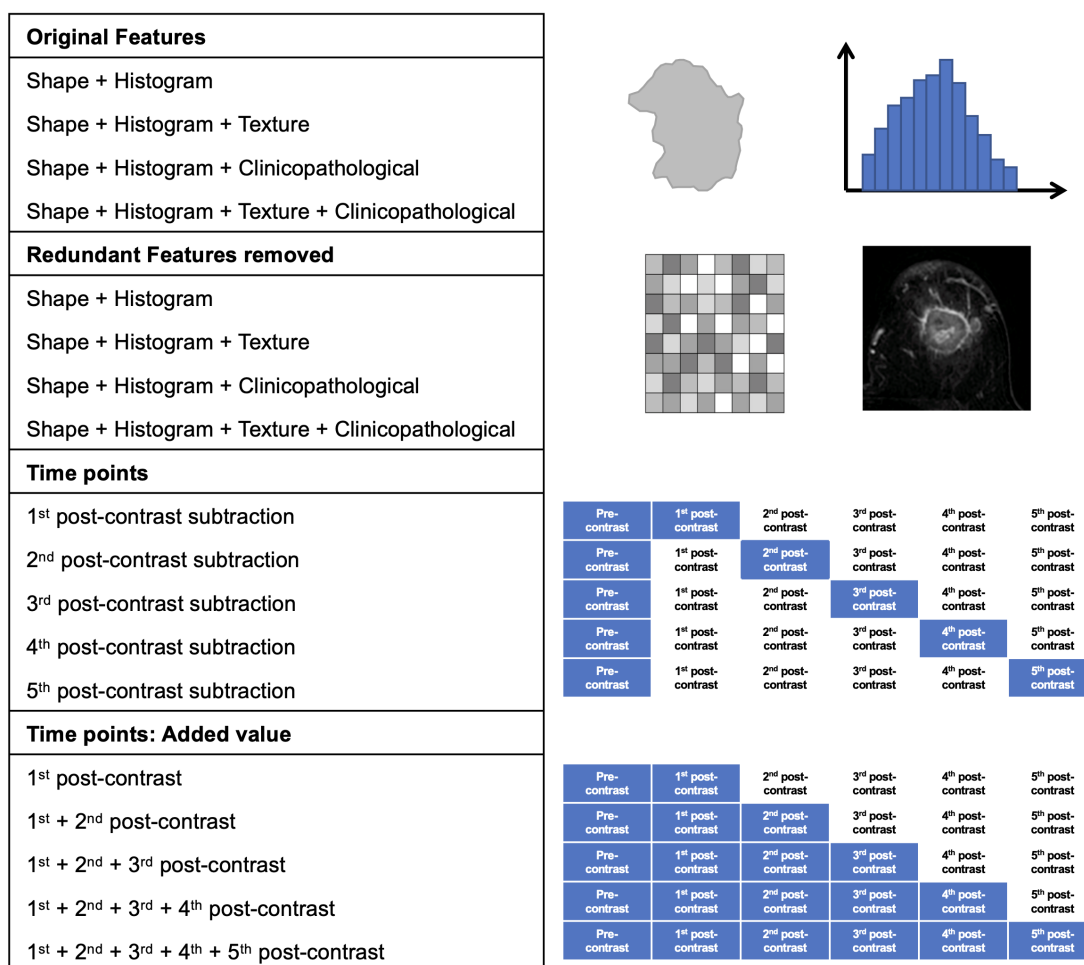


Figure. 6.2 Summary of the models trained using different combinations of features and post-contrast time points.

features were selected including 3 clinicopathological features, 4 shape features, and 12 histogram features. The model with the highest performance trained using S+H+T+P features selected only one feature, molecular subtype group. The test scores for models trained without clinicopathological features were low, suggesting overfitting. The 95% confidence intervals for the cross-validation scores were large for all feature sets. ROC curves for the model with the best performance for each feature set are shown in Figure 6.3B.

Feature reduction

The correlation matrix for all radiomics features is shown in Figure 6.4. In the S+H feature set, 60 strongly correlated features were removed including 4 shape features and 56 histogram features. In the S+H+T feature set, 284 strongly correlated features were removed including

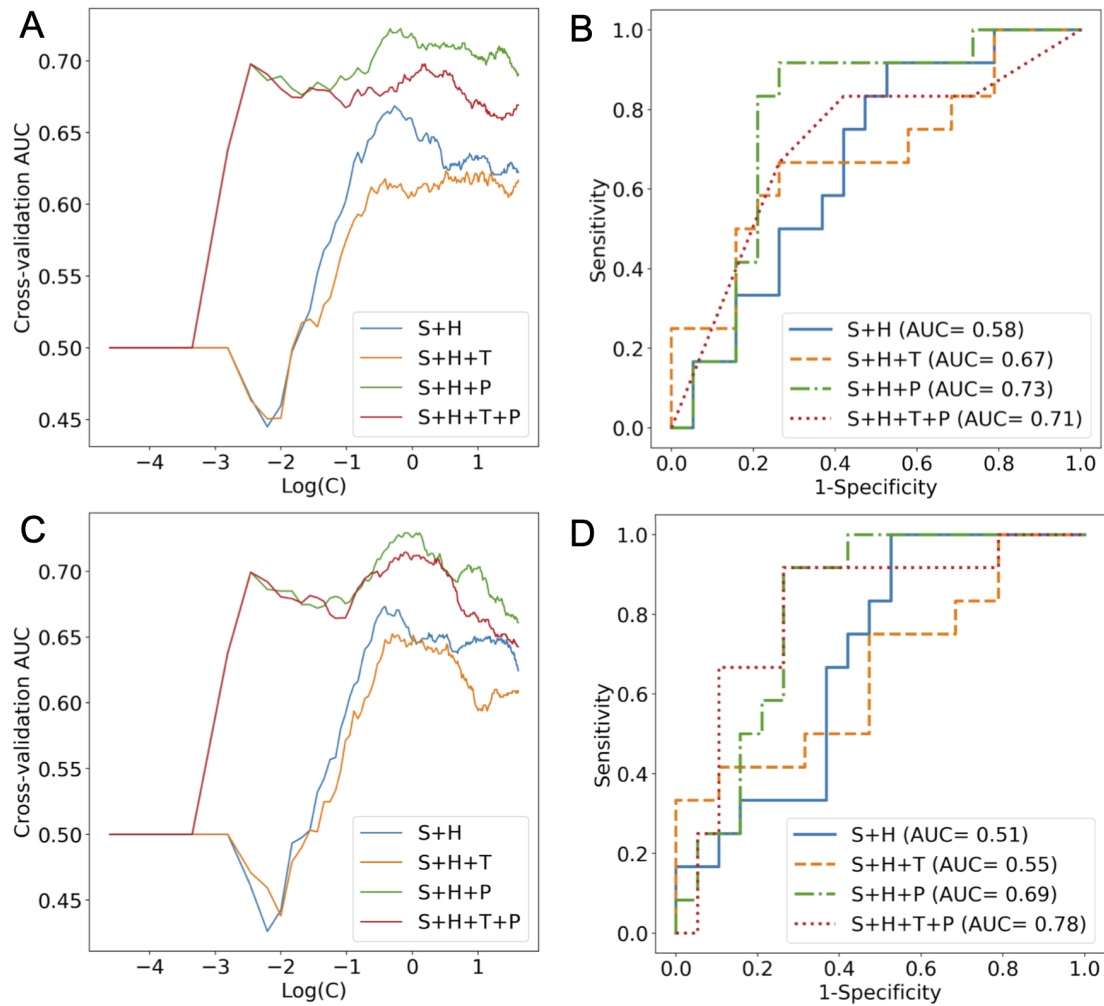


Figure 6.3 Cross-validations scores for models trained using A) all feature sets and C) all feature sets with redundant features removed with varied regularisation, and receiver operating characteristic (ROC) curves for models trained using B) all feature sets and D) all feature sets with redundant features removed. Area under the curve (AUC) in the test set is given for each feature set.

4 shape features, 60 histogram features, 87 GLCM features, 75 GLRLM features, and 58 GLSZM features. The highest proportion of redundant features was found in the GLRLM group (93.8% of features removed). No clinicopathological features were significantly correlated with any other feature.

Figure 6.3C shows the change in cross-validation score for each reduced feature set with regularisation parameter, C. The trends were similar to the models trained using the unreduced feature sets, though the models were more unstable with increasing numbers of features included. As before, the models trained using feature sets including clinicopathological

Table 6.3 Performance of models trained on each of the feature sets with 95% confidence intervals, and the number of features selected by the models.

	CV AUC (95% CI)	Test AUC (95% CI)	N features selected (Features available)
Shape + histogram	0.67 (0.50 - 0.93)	0.58 (0.50 - 0.75)	19 (104)
Shape + histogram + texture	0.62 (0.55 - 0.97)	0.67 (0.48 - 0.75)	47 (384)
Shape + histogram + pathology	0.72 (0.55 - 0.96)	0.73 (0.57 - 0.83)	19 (108)
Shape + histogram + texture + pathology	0.70 (0.44 - 0.92)	0.71 (0.53 - 0.78)	1 (388)
Redundant Features Removed			
Shape + histogram	0.67 (0.49 - 0.93)	0.51 (0.51 - 0.77)	14 (44)
Shape + histogram + texture	0.65 (0.54 - 0.96)	0.55 (0.49 - 0.76)	30 (100)
Shape + histogram + pathology	0.73 (0.55 - 0.96)	0.69 (0.57 - 0.85)	17 (48)
Shape + histogram + texture + pathology	0.71 (0.58 - 0.97)	0.78 (0.56 - 0.85)	30 (104)

CV: cross-validation, CI: confidence intervals, AUC: area under the curve

features had a similar peak in cross-validation score at 1 feature (molecular subtype group) with the same performance (cross-validation score 0.70), but both achieved higher cross-validation scores with additional features included in the model. The test scores for models trained without clinicopathological features were low.

The models trained on feature sets with redundant features removed achieved a higher cross-validation score but a worse test score than those trained on the full feature sets (Table 6.3). The highest cross-validation score was achieved using a model trained on S+H+P features, however the highest test score was achieved using a model trained on S+H+T+P features. The 95% confidence intervals for the cross-validation scores were large for all feature sets. ROC curves for the model with the best performance for each reduced feature set are shown in Figure 6.3D.

Feature selection

Features selected by all the models are given in Appendix 2. 7 features were selected by all models except the S+H+T+P model: 3 shape features (sphericity, maximum 2D diameter columns, and least axis length) and 4 histogram features (1st post-contrast kurtosis, 2nd

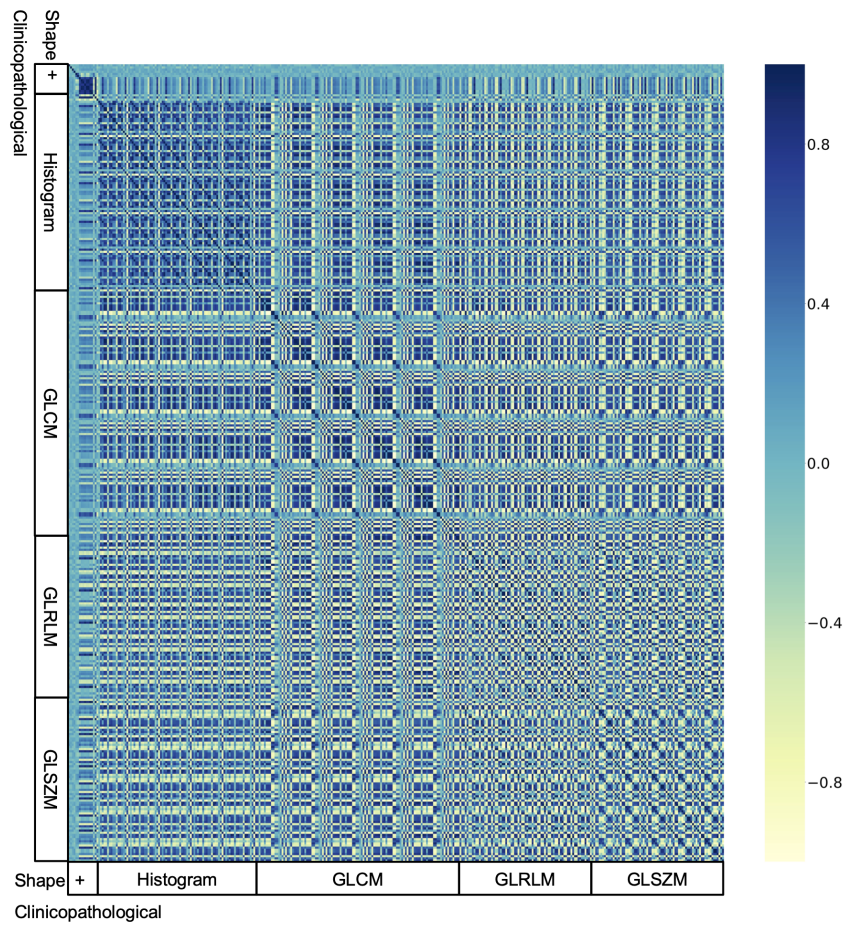


Figure. 6.4 The correlation matrix for all 388 features using Spearman's rank correlation coefficient.

post-contrast skewness, 3rd post-contrast kurtosis, and 5th post-contrast kurtosis), shown in Figure 6.5. For models trained on feature sets that included texture features, 7 texture features were selected by all models except the S+H+T+P model: 3 GLCM features (1st post-contrast informational measure of correlation (Imc2), 4th post-contrast maximal correlation coefficient (MCC), and 5th post-contrast MCC), 1 GLRLM feature (3rd post-contrast long run low grey level emphasis), and 3 GLSZM features (3rd post-contrast large area high grey level emphasis, 4th post-contrast small area emphasis, and 5th post-contrast size zone non-uniformity normalised).

The models trained using just shape and histogram features (S+H and S+H with redundant features removed) selected 12 of the same features (out of 19 and 14 total features selected, respectively). The models trained using feature sets including texture features (S+H+T and S+H+T with redundant features removed) selected 24 of the same features (out of 47 and 30

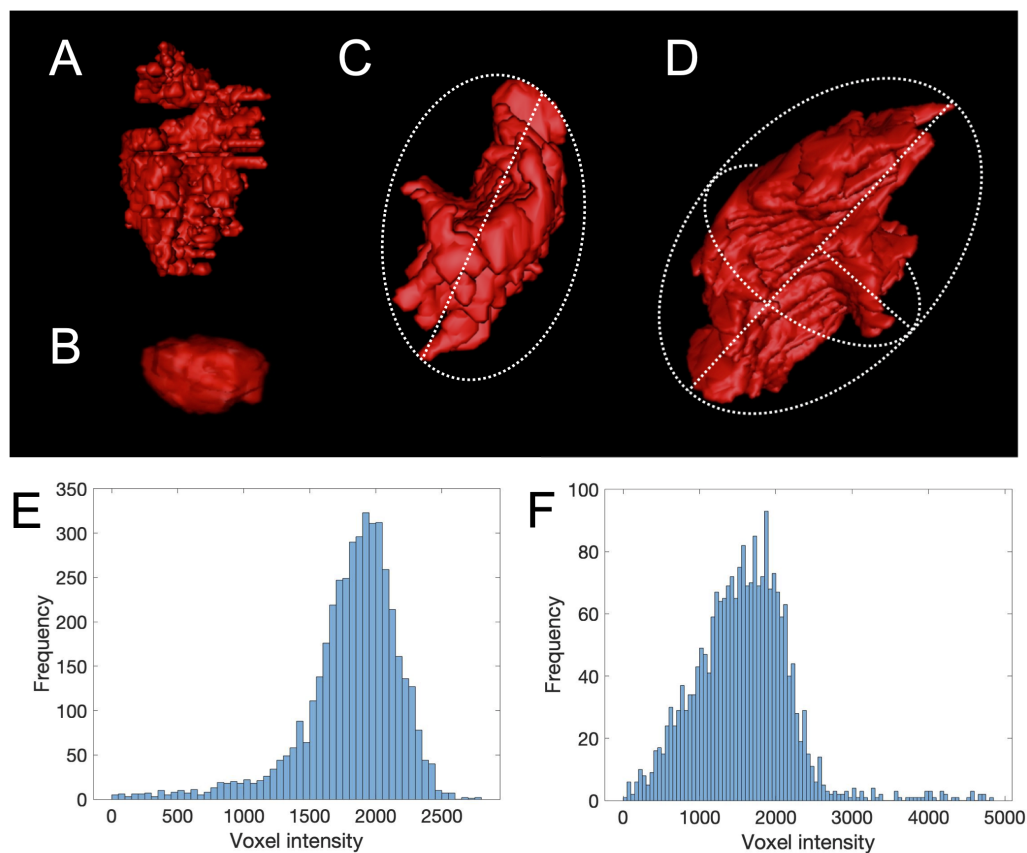


Figure. 6.5 Visualisation of shape and histogram features commonly selected by the models. The A) least and B) most spherical tumours (sphericity = 0.20 and 0.76, respectively). C) A representation of maximum 2D diameter column, which gives the largest pairwise Euclidean distance between tumour surface mesh vertices in the coronal plane. D) The least axis length is the smallest axis length of the ellipsoid that encloses the three-dimensional tumour ROI. E) The histogram of voxel values in the 2nd post-contrast subtraction image with the highest skewness (-1.64). Skewness measures the asymmetry of the distribution of voxel values about the mean value. F) The histogram of voxel values in the 1st post-contrast subtraction image with the highest kurtosis (5.72). A distribution of voxel values with a high kurtosis implies that the majority of the voxels are concentrated in the tails of the distribution as opposed to around the mean value.

total features selected, respectively) of which 13 were texture features. The models trained using shape, histogram, and clinicopathological features (S+H+P and S+H+P with redundant features removed) selected 11 of the same features (out of 19 and 17 features, respectively), of which 3 were clinicopathological features (age, grade, and molecular subtype group). The

models trained using all types of features (S+H+T+P and S+H+T+P with redundant features removed) selected only one feature that was the same (molecular subtype group).

After Benjamini-Hochberg correction, no shape, histogram, or texture features were statistically significantly different between pCR and non-pCR groups. There was a statistically significant difference in pCR and non-pCR groups based on molecular subtype ($p < 0.001$) and grade ($p = 0.01$) but not histopathological type ($p = 0.24$) or age ($p = 0.16$).

Comparison of post-contrast time points

Figure 6.6A shows the change in cross-validation score for models trained using image-derived features from each time point individually with regularisation parameter, C. All models reached a maximum cross-validation score then performance decreased with additional features included in the model. The cross-validation scores, test scores, and number of features selected by models trained using each time point are shown in Table 6.4. The highest cross-validation score was achieved using features from only the 3rd post-contrast time point and the highest test score was achieved using features from only the 1st post-contrast time point. The 95% confidence intervals for the cross-validation scores were large for all time points. ROC curves for the model with the best performance for each time point are shown in Figure 6.6B.

Figure 6.6C shows the change in cross-validation score for models trained using image-derived features from each time point individually plus clinicopathological features with regularisation parameter, C. For the 1st, 4th, and 5th post-contrast time points, models reached a maximum cross-validation score using only one feature (molecular subtype group) then performance decreased with additional features included in the model. For the 2nd and 3rd post-contrast time points, the cross-validation score improved with additional features included in the model up to a maximum at 5 features and 26 features, respectively, before performance decreased with additional features included. The cross-validation scores, test scores, and number of features selected by models trained using each time point are shown in Table 6.4. The highest cross-validation score was again achieved using features from only the 3rd post-contrast time point and the highest test score was achieved using only molecular subtype group (1st, 4th, and 5th post-contrast time points). The 95% confidence intervals for the cross-validation scores were large for all time points. ROC curves for the model with the best performance for each time point are shown in Figure 6.6D.

Features selected by all the models are given in Appendix 2. For models trained using only image-derived features, 8 features were selected by the models at every time point: 4 shape features (flatness, least axis length, maximum 2D diameter column, and sphericity), 1 histogram feature (skewness), and 3 texture features (correlation, MCC, and large area

low grey level emphasis). Kurtosis was selected by all models except the model trained using features from the 1st post-contrast time point. For models trained using image-derived features and clinicopathological features, molecular subtype group was the only feature selected at all post-contrast time points.

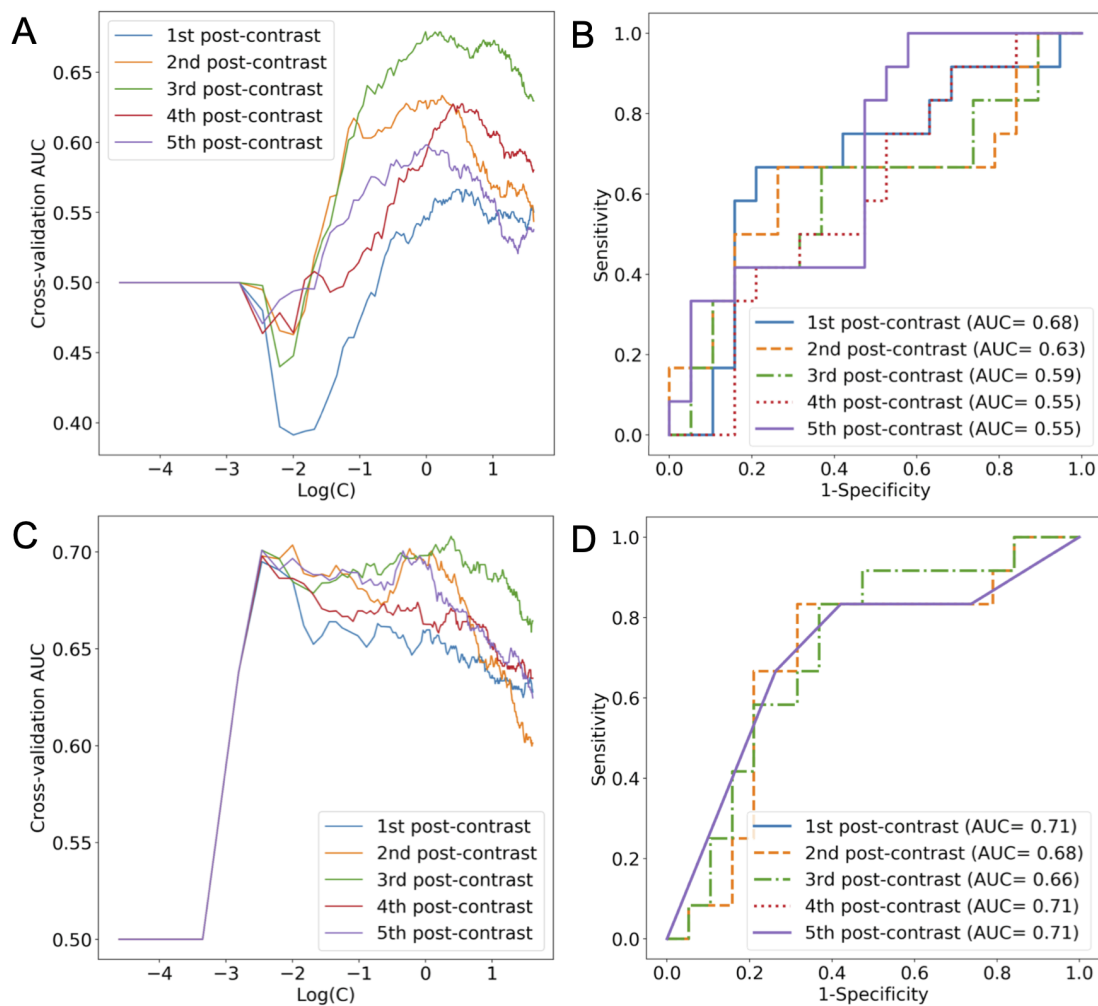


Figure. 6.6 Cross-validations scores for models trained using each time point individually using A) shape, histogram, and texture features and C) shape, histogram, texture, and clinicopathological features with varied regularisation, and receiver operating characteristic (ROC) curves for models trained using B) shape, histogram, and texture features and D) shape, histogram, texture, and clinicopathological features. Area under the curve (AUC) in the test set is given for each time point.

The cross-validation scores, test scores, and number of features selected by models trained using combinations of features from different time points are shown in Table 6.5. The highest cross-validation score was achieved using features from the full dynamic series (S+H+P with

Table 6.4 Performance of models trained using features from each time point with 95% confidence intervals, and the number of features selected by the models.

Time point	CV AUC (95% CI)	Test AUC (95% CI)	N features selected (Features available)
Image-derived features			
1st post-contrast	0.57 (0.47 - 0.92)	0.68 (0.47 - 0.77)	26 (88)
2nd post-contrast	0.63 (0.49 - 0.93)	0.63 (0.43 - 0.74)	19 (88)
3rd post-contrast	0.68 (0.50 - 0.93)	0.59 (0.49 - 0.75)	19 (88)
4th post-contrast	0.63 (0.51 - 0.94)	0.55 (0.46 - 0.73)	26 (88)
5th post-contrast	0.60 (0.46 - 0.91)	0.55 (0.50 - 0.80)	18 (88)
Image-derived features and clinicopathological features			
1st post-contrast	0.70 (0.44 - 0.91)	0.71 (0.54 - 0.79)	1 (92)
2nd post-contrast	0.70 (0.49 - 0.94)	0.68 (0.51 - 0.79)	5 (92)
3rd post-contrast	0.71 (0.55 - 0.96)	0.66 (0.52 - 0.81)	26 (92)
4th post-contrast	0.70 (0.44 - 0.92)	0.71 (0.53 - 0.79)	1 (92)
5th post-contrast	0.70 (0.44 - 0.92)	0.71 (0.58 - 0.79)	1 (92)

CV: cross-validation, CI: confidence intervals, AUC: area under the curve

correlated features removed). The models trained using features from each combination of time points achieved the same cross-validation score. From the model trained using only features from the 1st post-contrast time point which achieved a cross-validation score of 0.70 using only molecular subtype group, there was no additional improvement in performance with features from each additional time point added. The features selected by the models are given in Appendix 2.

6.4 Discussion

In this study, we demonstrated that machine learning models trained on shape, histogram, and texture features derived from pre-treatment MRI are able to predict pathological complete

Table 6.5 Performance of models trained using features combinations of time points with 95% confidence intervals, and the number of features selected by the models.

Time point	CV AUC (95% CI)	Test AUC (95% CI)	N features selected (Features available)
1st post-contrast	0.70 (0.44 - 0.91)	0.71 (0.54 - 0.79)	1 (92)
1st + 2nd post-contrast	0.70 (0.49 - 0.93)	0.68 (0.52 - 0.80)	5 (180)
1st + 2nd + 3rd post-contrast	0.70 (0.46 - 0.93)	0.68 (0.52 - 0.79)	6 (268)
1st + 2nd + 3rd + 4th post-contrast	0.70 (0.44 - 0.93)	0.71 (0.50 - 0.77)	1 (356)
All time points (S+H+P with correlated features removed)	0.73 (0.55 - 0.96)	0.69 (0.57 - 0.85)	17 (48)

CV: cross-validation, CI: confidence intervals, AUC: area under the curve, S: shape, H: histogram, P: clinicopathological

response to neoadjuvant chemotherapy in breast cancer with a moderate accuracy. The addition of clinicopathological variables improved the performance of the models, though a high accuracy was achieved using molecular subtype group independently without the use of image-derived features. While a number of studies have used logistic regression classification to predict pCR from pre-treatment MRI with a similar performance [141, 303, 304], this is the first study to evaluate the stability of the features selected using logistic regression with the LASSO and to interrogate the ability of the LASSO to handle collinearity in the feature set.

The training of machine learning models differs from traditional regression where the goal is to understand the impact of independent features on a dependent variable; this is not possible for feature sets with strong multicollinearity as is the case with a large set of radiomics features, particularly texture features. Logistic regression classification models using the LASSO are appealing due to the automatic variable selection which offers a sparse solution compared to other regression techniques [300, 305]. However, Zou and Hastie showed that for a group of features among which the pairwise correlations are very high, the regression coefficients of the variables tend to be equal and therefore LASSO tends to select one of the features randomly [306]. In this study, redundant strongly correlated features were removed before model training, aiming to improve prediction and feature selection. While the cross-validation scores overall improved after feature reduction, the lower test scores suggest more overfitting, though the features selected were similar to those selected by models trained on the full feature sets with there being a stable subset of selected features. Strong collinearity

in the feature set did not affect the stability of the performance of the models, in fact removing correlated features made the cross-validation performance more unstable for different degrees of regularisation. The elastic net is another regularisation technique proposed by Zou and Hastie to better handle the selection of strongly correlated features [306]. This technique should be used when the number of features is much greater than the number of data points ($p \gg n$), as is the case for radiomics studies using a greater number of radiomics features derived from multiparametric MRI but was not the case in this study (388 features and 152 patients). Furthermore, if $p > n$, LASSO can at most select n features before it saturates.

The low test scores observed in this study can be attributed to the large number of features selected by each model. The 'one in ten' rule is used as a rule of thumb for how many predictors can be fitted reliably in a regression [307, 308] and suggests that in this study, for a training set with 48 patients achieving pCR, a model that selects more than 5 features will likely overfit on the training data and will not perform well on test data. Given the large number of features selected by each model, they cannot be interpreted as a radiomic signature. However, there was a stable subset of 7 features that were selected by all models trained using features from all time points. Of the 3 shape features selected by the models, sphericity has been shown to be associated with pCR [309] and least axis length and maximum 2D diameter column are three-dimensional measures of tumour size, which have been shown to be more accurate in predicting tumour response than uni- or bi-dimensional measures of tumour size [112]. Of the histogram features selected, the kurtosis [146, 310, 311] and skewness [140, 312] of the enhancement of tumour voxels are well established parameters associated with tumour response. This demonstrates the ability of logistic regression classification using the LASSO to identify key features associated with response. Molecular subtype group had a high predictive performance when used independently, suggesting that pCR could be predicted with a high accuracy from just a core biopsy. However, the addition of imaging features in feature sets that included molecular subtype generally incrementally improved the performance of models. While some studies have found an improvement in performance when separating patients into groups by subtype [141, 313] or achieved a high predictive accuracy in a study of only triple negative breast cancers [119, 304], there were not enough cancers in each subtype group in this study to train machine learning models on each group independently. For models trained using feature sets including texture features, texture features make up a large proportion of features selected. The addition of texture features to feature sets generally did not improve the performance of models, and the feature selection was less stable given that a smaller proportion of the same features were selected by different models. The model with the highest cross-validation score in this study was trained on a feature set that did not include any texture features. Given the low test scores observed

this study, the models require validation on external test data sets to assess their ability to generalise to a larger target population.

Abbreviated MRI protocols are gaining widespread attention in the detection and diagnosis of breast cancer, where only one pre-contrast and one post-contrast time point are acquired. To date there have been no studies that have investigated an abbreviated protocol for the assessment of tumour response to chemotherapy. In this study, the model trained on image-derived features from the 1st post-contrast time point had a poor performance, and when clinicopathological features were included in the model, the performance improved but no image-derived features were included in the model. When comparing time points used independently, the model trained using features from the 3rd post-contrast time point achieved the best cross-validation score, suggesting that more than one post-contrast time point is necessary in a dynamic protocol for the prediction of pCR. The highest cross-validation score of any model was achieved using features from all time points. However this was only an incremental improvement from the use of molecular subtype as a univariate predictor of pCR, and this model had a lower test score.

When using features extracted from each post-contrast time point individually to train models, mostly shape and texture features were similarly selected. While tumour volume was not selected as a predictive feature by the models, functional tumour volume (calculated from the sum of voxels meeting a threshold for percentage enhancement using pre-contrast, early post-contrast, and late post-contrast images) has been shown to be a strong predictor of pCR and recurrence-free survival [113, 114]. The functional tumour volume could be used to generate a more sophisticated ROI from which to extract radiomics features that better encapsulates the tumour voxels of interest. Though the use each of the post-contrast time point individually was useful for the assessment of feature selection, it may be more useful from a clinical perspective to compare features extracted from quantitative maps of heuristic parameters that measure changes in enhancement between the post-contrast time points. A study by Loo et al. found that parameters derived from the early (10s after contrast injection) and late (450s after contrast injection) enhancement of tumours were associated with response to therapy [314]. The 1st post-contrast subtraction in this study captures the enhancement of voxels 1 minute after contrast injection; a time point acquired sooner after contrast injection may better capture rapidly enhancing voxels and demonstrate increased vascularity which is related to tumour aggressiveness. Parameters of wash-out, which measures the difference between peak enhancement and late enhancement, have been shown to be associated with tumour response [315–318] and could provide a more informative measure of contrast kinetics than using late post-contrast time points independently.

This study had some limitations. First, the ROIs were delineated by three radiologists independently. A study by Granzier et al. found a high spatial overlap (mean dice score 0.81) for ROIs delineated by observers of varying expertise, but only 32.8% of radiomics features extracted were found to be robust between observers [319]. As the radiomics features will be affected by the ROI definition, ROIs should be automatically segmented or delineated in consensus. Second, only logistic regression classification models were trained in this study. Studies have achieved a higher accuracy using other machine learning algorithms for radiomics analysis such as support vector machines [141, 313], random forests [143, 320] and XGB boost [321]. The performance of these models trained on the different feature sets should be compared. Third, only features extracted from DCE-MRI were used. Additional features extracted from other MRI sequences such as T₂-weighted images and diffusion-weighted images could be included in the analysis, though increasing the number of features significantly (such that $p \gg n$) may be problematic when using the LASSO. Fourth, the proportions of patients that achieved pCR and the proportion of patients with triple negative breast cancer were higher than average, and therefore this patient population may not be representative of a general breast cancer population. Finally, the patients included in this study received a range of types of chemotherapy (information not provided). The type and duration of chemotherapy regimen should in future be included as a feature in the models, as this will affect the rate of pCR.

In conclusion, while machine learning models trained using radiomics features derived from pre-treatment DCE-MRI were able to predict pCR with a moderate accuracy, tumour molecular subtype used independently was able to predict pCR with a similar accuracy. Logistic regression classification using the LASSO was able to handle collinearity in a feature set and a stable set of predictive features of response were selected. Although abbreviated protocols are increasingly used in the detection of breast cancer, this study has shown that a full dynamic series provides the most accurate prediction of pCR. Further work is required to validate these results on external test data sets.

Chapter 7

Future work and conclusions

7.1 Contributions to knowledge

This thesis has investigated the use of magnetic resonance imaging (MRI) in the detection of breast cancer and the prediction of pathological complete response to chemotherapy (pCR). The major contributions to knowledge of this thesis include a thorough systematic review and meta-analysis of the literature of diffusion-weighted imaging (DWI) techniques used in the detection of breast cancer, the optimisation and implementation of a multi-shot DWI protocol that improves upon the image quality of standard DWI, a comparative meta-analysis of abbreviated and full diagnostic protocol MRI in the detection of breast cancer, and the assessment of the performance of machine learning models trained using radiomic features derived from dynamic contrast-enhanced MRI (DCE-MRI) in the prediction of pCR.

The systematic review and meta-analysis presented in chapter 3 has demonstrated the comparable performance of DWI and DCE-MRI in the differential diagnosis of breast cancer. As non-contrast alternatives to DCE-MRI are increasingly sought out, this work contributes to the increasing evidence of DWI as a clinically equivalent diagnostic technique. This work also established that there is no reliable evidence to suggest that advanced DWI techniques offer superior clinical value in the differential diagnosis of breast lesions compared to standard assessment using the apparent diffusion coefficient (ADC). The conclusions of this study have been cited by the international DWI working group of the European Society of Breast Imaging in their consensus and mission statement [322].

The optimisation and validation of a DWI protocol using multiplexed sensitivity encoding (MUSE) in chapter 4 has demonstrated the superior image quality of multi-shot DWI compared to standard DWI acquired using single-shot echo-planar imaging (ss-EPI) by improving spatial resolution and reducing blurring and geometric distortion. As a result of this study, MUSE-DWI has replaced ss-EPI-DWI as part of the routine clinical breast MRI protocol

at Addenbrooke's hospital. This work will contribute to a growing literature showing the potential of advanced DWI acquisition techniques to increase the clinical utility of DWI.

The meta-analysis presented in chapter 5 has demonstrated the comparable diagnostic performance of abbreviated and full diagnostic MRI protocols in the detection of breast cancer. By undertaking a separate pooling of accuracy measures for studies reporting screening and enriched patient cohorts, this work has highlighted the difference in the performance of abbreviated protocols investigated in cohorts with a different prevalence of cancers. Given the increasing number of studies in the literature reporting abbreviated protocols, this work has provided a necessary interrogation of abbreviated MRI study methodology and emphasised the need for prospective trials to be undertaken in a large screening population with sufficient statistical power to demonstrate diagnostic equivalence to full diagnostic protocol MRI.

The radiomics analysis presented in chapter 6 has contributed to the evidence that radiomics features derived from pre-treatment DCE-MRI can be informative in the prediction of pathological complete response to chemotherapy. As well as confirming that machine learning techniques can be used to identify well established markers of response to treatment, namely immunohistochemical subtype, morphological features, and measures of irregular enhancement, this work has interrogated the ability of logistic regression classification using the least absolute shrinkage and selection operator (LASSO) to handle collinearity in a feature set and demonstrated the stability of feature selection. While the cross-validation performance of machine learning models trained using these features was moderate, it should be noted that the poor performance of the models tested on an internal hold-out validation dataset suggests the generalisability of these models to external test data may be limited. By comparing the features selected at different time points in the dynamic series, this work highlights a gap in the literature for the comparison of abbreviated and full dynamic protocols in the assessment of response to therapy.

7.2 Future work

Since the publication of the meta-analysis presented in Chapter 3, other advanced DWI techniques are increasingly investigated in the breast. A number of studies have investigated diffusion kurtosis imaging (DKI) (Section 1.2.3), which allows for a measurement of the non-Gaussianity of diffusion without the need for definition of tissue compartments or complex model fitting. Given the high diagnostic performance of DKI parameters [323–325], this technique should be systematically compared with other DWI models. While advanced DWI models offer limited additional improvement in cancer diagnosis compared to ADC assessment, in a research setting they can provide functional information about the cellular

and perfusion properties of breast tumours. One such application is sodium MRI (^{23}Na -MRI), which was investigated alongside this thesis (Appendix 3) and is ongoing research in the department. ^{23}Na -MRI can measure intra- and extra-cellular sodium concentrations in breast tumours, with DWI providing complementary physiological information about tumour cellularity and the integrity of cellular membranes and structures.

Currently, DWI is often used as an adjunct to DCE-MRI to improve specificity and reduce false positive results and recall rates for the screening of high-risk women. However, the high cost of MRI and concerns of gadolinium retention with repeated administration of contrast agent limit the widespread use of DCE-MRI as a screening tool for moderate-risk women. DWI has been shown to achieve a diagnostic accuracy similar to DCE-MRI without the need for intravenous contrast agent, though significant improvements in image quality and spatial resolution are required before a non-contrast DWI screening protocol can be implemented clinically. While the MUSE-DWI technique has been shown to improve image quality, further improvements come at the expense of increased scan time with additional numbers of 'shots'. Studies have demonstrated a reduction in acquisition time of multi-shot techniques by implementing simultaneous multi-slice techniques [265], however these techniques often require breast coils with a higher number of receiver coils (≥ 16) which are not widely used for routine clinical scanning.

Future work should focus on delivering high quality DWI in a clinically feasible acquisition time that could be used for screening. MUSE is just one of a number of promising techniques including readout-segmented EPI [254, 257, 326], partial-FOV DWI [327, 328], axially reformatted simultaneous multi-slice imaging [326], and double echo steady state (DESS) DWI [329], that attempt to overcome the limitations of conventional DWI acquisition in the breast. In order to ascertain which is the best technique, a head-to-head comparison is required. A spatial resolution of close to 1mm to distinguish lesions $<10\text{mm}$ (as the detection of these small lesions has been shown to reduce breast cancer mortality [87]) acquired within a 5-minute examination time is required for DWI to be clinically feasible as a screening technique. Furthermore, these techniques require validation in large patient cohorts that are representative of a true screening population. To date, preliminary studies lack the statistical power required to prove equivalent diagnostic performance of these techniques with DCE-MRI.

MRI as a supplemental screening modality is of interest in cohorts more at risk than the general population (such as gene carriers and those with a previous history of breast cancer) and in populations where existing screening techniques are insufficient, such as women with dense breasts. The Breast Screening- Risk Adaptive Imaging for Density (BRAID) trial is large prospective multi-centre study currently underway led by the University of Cambridge

[330], comparing supplemental imaging using abbreviated MRI, automated whole breast ultrasound (ABUS), and contrast enhanced spectral mammography (CESM) with standard of care (no supplemental imaging) in women with dense breasts. This study will enable a measure of the diagnostic performance of abbreviated MRI in a target population of thousands of women.

Radiomics and machine learning are rapidly growing fields of breast MRI research given the advances in computational power and development of specialised programming packages (such as scikit-learn, PyTorch, and TensorFlow [302, 331–333]) for the application of machine learning and deep learning techniques. While there were issues related to access of images during the course of this work, the dataset used for radiomics analysis also includes T_2 -weighted images and diffusion-weighted images acquired at multiple b-values, as well as B_1 and T_1 mapping sequences to generate quantitative T_1 maps. The same set of radiomics features should be extracted from these images and machine learning models trained on a larger and more diverse set of features that make the most of this multiparametric MRI dataset. High resolution DWI such as MUSE-DWI would also be ideal for the extraction of texture features to better measure the intratumoural heterogeneity of diffusion. The features selected by machine learning models trained on DCE-MRI, DWI, and both DCE-MRI and DWI features should be compared. These features could better our understanding of imaging biomarkers of tumour aggressiveness and inform the use of DWI for screening, where the use of informative radiomics features may improve sensitivity.

While we have shown a reasonable accuracy of predictive pCR models on hold-out test data within our study, this work requires validation on external test datasets. The ISPY1 (ACRIN 6657) trial is a publicly available multicentre breast MRI dataset with lesion segmentations, pathological information, and pCR outcomes that would be suitable for this purpose [334]. Multicentre data sets are particularly crucial to compile test data that is fully representative of a population of women with breast cancer. The limited availability of such data sets is driving the creation of imaging repositories for testing of machine learning algorithms, though issues surrounding the reproducibility of radiomics features arising from the variation in image acquisition protocols must be addressed [335, 336]. The Image Biomarker Standardisation Initiative (IBSI) has been formed to address the lack of consistency in radiomics analyses by establishing mathematical definitions for commonly used radiomics features, suggested methodology for image pre-processing and feature calculation, reference datasets for calibration of software implementations (including digital phantoms), and standardised reporting guidelines for studies involving radiomic analyses [337]. The next stage of this work should ensure that the radiomics analysis workflow adheres to these guidelines.

In contrast to use of radiomics features, convolutional neural networks (CNNs) are increasingly used in the prediction of pCR, eliminating the need for the choice of features and the time-consuming manual delineation of breast lesions by breast radiologists. Given the performance of three-dimensional shape features in the prediction of response, a three-dimensional network should be developed to predict pCR from pre-treatment DCE-MRI. This is a novel area of research given that neural networks using breast MRI to predict NACT response have been developed using network architecture with 2D inputs to date.

7.3 Conclusions

1. Diffusion-weighted imaging (DWI), intra-voxel incoherent motion (IVIM), and diffusion tensor imaging (DTI) are able to discriminate between malignant and benign breast lesions with a high accuracy comparable that of dynamic contrast-enhanced MRI (DCE-MRI).
2. IVIM is diagnostically comparable to the apparent diffusion coefficient (ADC). DTI achieves a higher diagnostic accuracy than the ADC. However, given that the number of studies that have investigated these techniques to date is limited, there is insufficient statistical power to conclude equivalent or superior performance.
3. ADC, IVIM and DTI lack standardisation in reported methodology and acquisition parameters.
4. Multi-shot echo-planar imaging using multiplexed sensitivity encoding (MUSE) improves the image quality of DWI compared to single-shot echo-planar imaging by significantly reducing geometric distortion and blurring.
5. The diagnostic performance of abbreviated MRI is equivalent to that of a full diagnostic protocol amongst enriched cohorts and was lower but not significantly different in a screening cohort.
6. The diagnostic performance of abbreviated MRI varies in cohorts with a different prevalence of cancers. Evidence of diagnostic equivalence from prospective trials with sufficient statistical power is required before abbreviated MRI can be confidently adopted into clinical practice.
7. Machine learning models trained using radiomics features derived from pre-treatment DCE-MRI can predict pathological complete response (pCR) with a moderate accuracy.

8. Logistic regression classification using the least absolute shrinkage and selection operator (LASSO) can be used to find a stable set of radiomics features that are predictive of response to chemotherapy.
9. The most commonly selected features in the prediction of pCR are molecular subtype group, three-dimensional shape features (least axis length, sphericity, and maximum 2D diameter column) and histogram features (skewness and kurtosis).
10. Using machine learning models, a full dynamic series enables the most accurate prediction of pCR.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424, (2018).
- [2] Cancer Research UK. Breast cancer statistics | Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer>, (2020).
- [3] M. G. Marmot, D. G. Altman, D. A. Cameron, et al. The benefits and harms of breast cancer screening: An independent review. *British Journal of Cancer*, 108(11):2205–2240, (2013).
- [4] A. Vourtsis and W. A. Berg. Breast density implications and supplemental screening. *European Radiology*, 29(4):1762–1777, (2019).
- [5] T. M. Kolb, J. Lichy, and J. H. Newhouse. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: An analysis of 27,825 patient evaluations. *Radiology*, 225(1):165–175, (2002).
- [6] J. O. P. Wanders, K. Holland, W. B. Veldhuis, et al. Volumetric breast density affects performance of digital screening mammography. *Breast Cancer Research and Treatment*, 162(1):95–103, (2017).
- [7] E. Warner. Screening BRCA1 and BRCA2 mutation carriers for breast cancer. *Cancers*, 10(12), (2018).
- [8] F. Sardanelli and F. Podo, editors. *Breast MRI for High-risk Screening*. Springer International Publishing, Cham, (2020).
- [9] C. Kuhl, S. Weigel, S. Schrading, et al. Prospective multicenter cohort study to refine Kuhl, C., Weigel, S., Schrading, S., Arand, B., Bieling, H., König, R., . . . Schild, H. H. (2010). Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of b. *Journal of Clinical Oncology*, 28(9):1450–1457, (2010).
- [10] L. Zhang, M. Tang, Z. Min, et al. Accuracy of combined dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection: a meta-analysis. *Acta Radiologica*, 57(6):651–660, (2016).

- [11] G. C. Wishart, E. M. Azzato, D. C. Greenberg, et al. PREDICT: A new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Research*, 12(1):R1, (2010).
- [12] B. C. Pestalozzi, D. Zahrieh, E. Mallon, et al. Distinct Clinical and Prognostic Features of Infiltrating Lobular Carcinoma of the Breast: Combined Results of 15 International Breast Cancer Study Group Clinical Trials. *Journal of Clinical Oncology*, 26(18):3006–3014, (2008).
- [13] G. Ciriello, M. L. Gatz, A. H. Beck, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, 163(2):506–519, (2015).
- [14] E. A. Rakha, J. S. Reis-Filho, F. Baehner, et al. Breast cancer prognostic classification in the molecular era: The role of histological grade. *Breast Cancer Research*, 12(4):207, (2010).
- [15] H. Pereira, S. E. Pinder, D. M. Sibbering, et al. Pathological prognostic factors in breast cancer. IV: Should you be a typer or a grader? A comparative study of two histological prognostic features in operable breast carcinoma. *Histopathology*, 27(3):219–226, (1995).
- [16] C. Elston and I. Ellis. pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19(5):403–410, (1991).
- [17] A. G. Waks and E. P. Winer. Breast Cancer Treatment: A Review. *JAMA - Journal of the American Medical Association*, 321(3):288–300, (2019).
- [18] B. E. Henderson and H. S. Feigelson. Hormonal carcinogenesis The key distinction between this 'cell proliferation' model. Technical Report 3, (2000).
- [19] W. Yue, J. P. Wang, Y. Li, et al. Tamoxifen versus aromatase inhibitors for breast cancer prevention. In *Clinical Cancer Research*, volume 11, (2005).
- [20] M. R. Boland, J. Ryan, E. Dunne, et al. Meta-analysis of the impact of progesterone receptor status on oncological outcomes in oestrogen receptor-positive breast cancer. *British Journal of Surgery*, 107(1):33–43, (2020).
- [21] D. Slamon, G. Clark, S. Wong, et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, 235(4785):177–182, (1987).
- [22] Z. Mitri, T. Constantine, and R. O'Regan. The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemotherapy Research and Practice*, 2012:1–7, (2012).
- [23] T. Sørli, C. M. Perou, R. Tibshirani, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10869–10874, (2001).

- [24] C. M. Perou, T. Sørile, M. B. Eisen, et al. Molecular portraits of human breast tumours. *Nature*, 406(6797):747–752, (2000).
- [25] C. Curtis, S. P. Shah, S. F. Chin, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403):346–352, (2012).
- [26] R. Verma, R. L. Bowen, S. E. Slater, F. Mihaimeed, and J. L. Jones. Pathological and epidemiological factors associated with advanced stage at diagnosis of breast cancer. *British medical bulletin*, 103(1):129–45, (2012).
- [27] S. Iborra and E. Stickeler. HER2-Orientated Therapy in Early and Metastatic Breast Cancer. *Breast Care*, 11(6):392–397, (2016).
- [28] F. Podo, L. M. Buydens, H. Degani, et al. Triple-negative breast cancer: Present challenges and new perspectives. *Molecular Oncology*, 4(3):209–229, (2010).
- [29] A. Bamias and M. A. Dimopoulos. Angiogenesis in human cancer: Implications in cancer therapy. *European Journal of Internal Medicine*, 14(8):459–469, (2003).
- [30] D. W. Siemann. The unique characteristics of tumor vasculature and preclinical evidence for its selective disruption by Tumor-Vascular Disrupting Agents. *Cancer Treatment Reviews*, 37(1):63–74, (2011).
- [31] A. Grubstein, M. Yepes, and R. Kiszonas. Magnetic resonance imaging of breast vascularity in medial versus lateral breast cancer. *European Journal of Radiology*, 75(2), (2010).
- [32] L. D. Buadu, J. Murakami, S. Murayama, et al. Breast lesions: Correlation of contrast medium enhancement patterns on MR images with histopathologic findings and tumor angiogenesis. *Radiology*, 200(3):639–649, (1996).
- [33] ACR. ACR BI-RADS ATLAS - Breast MRI. <https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/MRI-Reporting.pdf>, (2013).
- [34] C. K. Kuhl, P. Mielcareck, S. Klaschik, et al. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology*, 211(1):101–110, (1999).
- [35] P. S. Tofts, G. Brix, D. L. Buckley, et al. Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted MRI of a diffusable tracer: Standardized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 10(3):223–232, (1999).
- [36] S. Sinha and U. Sinha. Functional magnetic resonance of human breast tumors: Diffusion and perfusion imaging. In *Annals of the New York Academy of Sciences*, volume 980, pages 95–115. New York Academy of Sciences, (2002).
- [37] D. Le Bihan, J. F. Mangin, C. Poupon, et al. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–546, (2001).
- [38] Y. Tsushima, A. Takahashi-Taketomi, and K. Endo. Magnetic resonance (MR) differential diagnosis of breast tumors using apparent diffusion coefficient (ADC) on 1.5-T. *Journal of Magnetic Resonance Imaging*, 30(2):249–255, (2009).

- [39] Y. Guo, Y.-Q. Cai, Z.-L. Cai, et al. Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. *Journal of Magnetic Resonance Imaging*, 16(2):172–178, (2002).
- [40] M. Hatakenaka, H. Soeda, H. Yabuuchi, et al. Apparent diffusion coefficients of breast tumors: Clinical application. *Magnetic Resonance in Medical Sciences*, 7(1):23–29, (2008).
- [41] S. Y. Choi, Y.-W. Chang, H. J. Park, et al. Correlation of the apparent diffusion coefficient values on diffusion-weighted imaging with prognostic factors for breast cancer. *The British journal of radiology*, 85(1016):474–9, (2012).
- [42] M. Costantini, P. Belli, P. Rinaldi, et al. Diffusion-weighted imaging in breast cancer: Relationship between apparent diffusion coefficient and tumour aggressiveness. *Clinical Radiology*, 65(12):1005–1012, (2010).
- [43] A. A. K. A. Razek, G. Gaballa, A. Denewer, and N. Nada. Invasive ductal carcinoma: correlation of apparent diffusion coefficient value with pathological prognostic factors. *NMR in Biomedicine*, 23(6):619–623, (2010).
- [44] C. Marini, C. Iaconi, M. Giannelli, et al. Quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesion. *European Radiology*, 17(10):2646–2655, (2007).
- [45] R. Woodhams, S. Kakita, H. Hata, et al. Diffusion-weighted imaging of mucinous carcinoma of the breast: Evaluation of apparent diffusion coefficient and signal intensity in correlation with histologic findings. *American Journal of Roentgenology*, 193(1):260–266, (2009).
- [46] T. Bozkurt Bostan, G. Koç, G. Sezgin, et al. Value of Apparent Diffusion Coefficient Values in Differentiating Malignant and Benign Breast Lesions. *Balkan medical journal*, 33(3):294–300, (2016).
- [47] M. Moschetta, M. Telegrafo, L. Rella, et al. MR evaluation of breast lesions obtained by diffusion-weighted imaging with background body signal suppression (DWIBS) and correlations with histological findings. *Magnetic Resonance Imaging*, 32(6):605–609, (2014).
- [48] M. Telegrafo, L. Rella, A. A. Stabile Ianora, G. Angelelli, and M. Moschetta. Unenhanced breast MRI (STIR, T2-weighted TSE, DWIBS): An accurate and alternative strategy for detecting and differentiating breast lesions. *Magnetic Resonance Imaging*, 33(8):951–955, (2015).
- [49] D. Le Bihan, E. Breton, D. Lallemand, et al. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*, 168(2):497–505, (1988).
- [50] G. C. Baxter, M. J. Graves, F. J. Gilbert, and A. J. Patterson. A Meta-analysis of the Diagnostic Performance of Diffusion MRI for Breast Lesion Characterization. *Radiology*, 291(3):632–641, (2019).
- [51] M. Iima, M. Honda, E. E. Sigmund, et al. Diffusion MRI of the breast: Current status and future directions. *Journal of Magnetic Resonance Imaging*, (2019).

- [52] J. H. Jensen, J. A. Helpert, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53(6):1432–1440, (2005).
- [53] L. Nogueira, S. Brandão, E. Matos, et al. Application of the diffusion kurtosis model for the study of breast lesions. *European Radiology*, 24(6):1197–1203, (2014).
- [54] D. Wu, G. Li, J. Zhang, et al. Characterization of Breast Tumors Using Diffusion Kurtosis Imaging (DKI). *PLoS ONE*, 9(11):e113240, (2014).
- [55] K. M. Bennett, K. M. Schmainda, R. Bennett (Tong), et al. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magnetic Resonance in Medicine*, 50(4):727–734, (2003).
- [56] S. Suo, F. Cheng, M. Cao, et al. Multiparametric diffusion-weighted imaging in breast lesions: Association with pathologic diagnosis and prognostic factors. *Journal of Magnetic Resonance Imaging*, 46(3):740–750, (2017).
- [57] R. Bedair, A. N. Priest, A. J. Patterson, et al. Assessment of early treatment response to neoadjuvant chemotherapy in breast cancer using non-mono-exponential diffusion models: a feasibility study comparing the baseline and mid-treatment MRI examinations. *European Radiology*, 27(7):2726–2736, (2017).
- [58] K. Yamaguchi, T. Nakazono, R. Egashira, et al. Diagnostic Performance of Diffusion Tensor Imaging with Readout-segmented Echo-planar Imaging for Invasive Breast Cancer: Correlation of ADC and FA with Pathological Prognostic Markers. *Magnetic Resonance in Medical Sciences*, 16(3):245–252, (2017).
- [59] R. Jiang, X. Zeng, S. Sun, Z. Ma, and X. Wang. Assessing Detection, Discrimination, and Risk of Breast Cancer According to Anisotropy Parameters of Diffusion Tensor Imaging. *Medical science monitor : international medical journal of experimental and clinical research*, 22:1318–28, (2016).
- [60] O. Cakir, A. Arslan, N. Inan, et al. Comparison of the diagnostic performances of diffusion parameters in diffusion weighted imaging and diffusion tensor imaging of breast lesions. *European Journal of Radiology*, 82(12):e801–e806, (2013).
- [61] P. A. T. Baltzer, A. Schäfer, M. Dietzel, et al. Diffusion tensor magnetic resonance imaging of the breast: a pilot study. *European Radiology*, 21(1):1–10, (2011).
- [62] C. Onaygil, H. Kaya, M. U. Ugurlu, and E. Aribal. Diagnostic performance of diffusion tensor imaging parameters in breast cancer and correlation with the prognostic factors. *Journal of Magnetic Resonance Imaging*, 45(3):660–672, (2017).
- [63] E. Furman-Haran, E. Eyal, M. Shapiro-Feinberg, et al. Advantages and drawbacks of breast DTI. *European Journal of Radiology*, 81:S45–S47, (2012).
- [64] S. C. Partridge, A. Ziadloo, R. Murthy, et al. Diffusion tensor MRI: Preliminary anisotropy measures and mapping of breast tumors. *Journal of Magnetic Resonance Imaging*, 31(2):339–347, (2010).

- [65] P. Belli, E. Bufi, A. Bonatesta, et al. Unenhanced breast magnetic resonance imaging: detection of breast cancer. *European review for medical and pharmacological sciences*, 20(20):4220–4229, (2016).
- [66] T. Kazama, Y. Kuroki, M. Kikuchi, et al. Diffusion-weighted MRI as an adjunct to mammography in women under 50 years of age: An initial study. *Journal of Magnetic Resonance Imaging*, 36(1):139–144, (2012).
- [67] S. C. Partridge, N. Nissan, H. Rahbar, A. E. Kitsch, and E. E. Sigmund. Diffusion-weighted breast MRI: Clinical applications and emerging techniques. *Journal of Magnetic Resonance Imaging*, 45(2):337–355, (2017).
- [68] S. K. Plevritis, A. W. Kurian, B. M. Sigal, et al. Cost-effectiveness of screening BRCA1/2 mutation carriers with breast magnetic resonance imaging. *Journal of the American Medical Association*, 295(20):2374–2384, (2006).
- [69] N. F. Boyd, E. Huszti, O. Melnichouk, et al. Mammographic features associated with interval breast cancers in screening programs. *Breast Cancer Research*, 16(1):417, (2014).
- [70] C. K. Kuhl, S. Schrading, K. Strobel, et al. Abbreviated breast Magnetic Resonance Imaging (MRI): First postcontrast subtracted images and maximum-intensity projection - A novel approach to breast cancer screening with MRI. *Journal of Clinical Oncology*, 32(22):2304–2310, (2014).
- [71] L. J. Grimm, M. S. Soo, S. Yoon, et al. Abbreviated Screening Protocol for Breast MRI. A Feasibility Study. *Academic Radiology*, 22(9):1157–1162, (2015).
- [72] V. Dialani, I. Tseng, P. J. Slanetz, et al. Potential role of abbreviated MRI for breast cancer screening in an academic medical center. *The Breast Journal*, 25(4):604–611, (2019).
- [73] S.-Q. Chen, M. Huang, Y.-Y. Shen, C.-L. Liu, and C.-X. Xu. Application of Abbreviated Protocol of Magnetic Resonance Imaging for Breast Cancer Screening in Dense Breast Tissue. *Academic Radiology*, 24(3):316–320, (2017).
- [74] J. M. Lee, L. Ichikawa, E. Valencia, et al. Performance Benchmarks for Screening Breast MR Imaging in Community Practice. *Radiology*, 285(1):44–52, (2017).
- [75] R. M. Strigel, J. Rollenhagen, E. S. Burnside, et al. Screening Breast MRI Outcomes in Routine Clinical Practice: Comparison to BI-RADS Benchmarks. *Academic Radiology*, 24(4):411–417, (2017).
- [76] M. F. Bakker, S. V. de Lange, R. M. Pijnappel, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. *New England Journal of Medicine*, 381(22):2091–2102, (2019).
- [77] C. E. Comstock, C. Gatsonis, G. M. Newstead, et al. Comparison of Abbreviated Breast MRI vs Digital Breast Tomosynthesis for Breast Cancer Detection Among Women With Dense Breasts Undergoing Screening. *JAMA*, 323(8):746, (2020).

- [78] S. P. Weinstein, K. Korhonen, C. Cirelli, et al. Abbreviated Breast Magnetic Resonance Imaging for Supplemental Screening of Women With Dense Breasts and Average Risk. *Journal of Clinical Oncology*, page JCO.19.02198, (2020).
- [79] J. S. McDonald, C. H. Hunt, A. B. Kolbe, et al. Acute Adverse Events Following Gadolinium-based Contrast Agent Administration: A Single-Center Retrospective Study of 281 945 Injections. *Radiology*, 292(3):620–627, (2019).
- [80] T. Kanda, K. Ishii, H. Kawaguchi, K. Kitajima, and D. Takenaka. High signal intensity in the dentate nucleus and globus pallidus on unenhanced T1-weighted MR images: Relationship with increasing cumulative dose of a gadoliniumbased contrast material. *Radiology*, 270(3):834–841, (2014).
- [81] S. Bickelhaupt, F. B. Laun, J. Tesdorff, et al. Fast and noninvasive characterization of suspicious lesions detected at breast cancer X-ray screening: Capability of diffusion-weighted MR imaging with MIPs. *Radiology*, 278(3):689–697, (2016).
- [82] P. A. Baltzer, M. Benndorf, M. Dietzel, et al. Sensitivity and specificity of unenhanced MR mammography (DWI combined with T2-weighted TSE imaging, ueMRM) for the differentiation of mass lesions. *European Radiology*, 20(5):1101–1110, (2010).
- [83] A. Rotili, R. M. Trimboli, S. Penco, et al. Double reading of diffusion-weighted magnetic resonance imaging for breast cancer detection. *Breast Cancer Research and Treatment*, 180(1):111–120, (2020).
- [84] J. W. Kang, H. J. Shin, K. C. Shin, et al. Unenhanced magnetic resonance screening using fused diffusion-weighted imaging and maximum-intensity projection in patients with a personal history of breast cancer: role of fused DWI for postoperative screening. *Breast Cancer Research and Treatment*, 165(1):119–128, (2017).
- [85] H. Yabuuchi, Y. Matsuo, S. Sunami, et al. Detection of non-palpable breast cancer in asymptomatic women by using unenhanced diffusion-weighted and T2-weighted MR imaging: Comparison with mammography and dynamic contrast-enhanced MR imaging. *European Radiology*, 21(1):11–17, (2011).
- [86] Y. Bu, J. Xia, B. Joseph, et al. Non-contrast MRI for breast screening: preliminary study on detectability of benign and malignant lesions in women with dense breasts. *Breast Cancer Research and Treatment*, 177(3):629–639, (2019).
- [87] L. Tabár, S. W. Duffy, B. Vitak, H. H. Chen, and T. C. Prevost. The natural history of breast carcinoma: What have we learned from screening? *Cancer*, 86(3):449–462, (1999).
- [88] T. A. King and M. Morrow. Surgical issues in patients with breast cancer receiving neoadjuvant chemotherapy. *Nature Reviews Clinical Oncology*, 12(6):335–343, (2015).
- [89] R. Wuerstlein and N. Harbeck. Neoadjuvant Therapy for HER2-positive Breast Cancer. *Reviews on Recent Clinical Trials*, 12(2):81–92, (2017).
- [90] C. Liedtke, C. Mazouni, K. R. Hess, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *Journal of Clinical Oncology*, 26(8):1275–1281, (2008).

- [91] P. Fong, D. Boss, T. Yap, and A. Tutt. Supplementary Material: Inhibition of poly (ADP-ribose) polymerase in tumors from BRCA mutation carriers. *New England Journal of Medicine*, 361(2):123–134, (2009).
- [92] B. Sirohi, M. Arnedos, S. Popat, et al. Platinum-based chemotherapy in triple-negative breast cancer. *Annals of Oncology*, 19(11):1847–1852, (2008).
- [93] P. Chollet, S. Amat, H. Cure, et al. Prognostic significance of a complete pathological response after induction chemotherapy in operable breast cancer. *British Journal of Cancer*, 86(7):1041–1046, (2002).
- [94] P. Cortazar, L. Zhang, M. Untch, et al. Pathological complete response and long-term clinical benefit in breast cancer: The CTNeoBC pooled analysis. *The Lancet*, 384(9938):164–172, (2014).
- [95] M. Teshome and K. K. Hunt. Neoadjuvant therapy in the treatment of breast cancer. *Surgical Oncology Clinics of North America*, 23(3):505–523, (2014).
- [96] N. Houssami, S. Ciatto, P. Macaskill, et al. Accuracy and surgical impact of magnetic resonance imaging in breast cancer staging: Systematic review and meta-analysis in detection of multifocal and multicentric cancer. *Journal of Clinical Oncology*, 26(19):3248–3258, (2008).
- [97] G. Mariscotti, N. Houssami, M. Durando, et al. Accuracy of mammography, digital breast tomosynthesis, ultrasound and MR imaging in preoperative assessment of breast cancer. *Anticancer Research*, 34(3):1219–1225, (2014).
- [98] M. B. Lobbes, R. Prevos, M. Smidt, et al. The role of magnetic resonance imaging in assessing residual disease and pathologic complete response in breast cancer patients receiving neoadjuvant chemotherapy: A systematic review. *Insights into Imaging*, 4(2):163–175, (2013).
- [99] Y. Yuan, X. S. Chen, S. Y. Liu, and K. W. Shen. Accuracy of MRI in prediction of pathologic complete remission in breast cancer after preoperative therapy: A meta-analysis. *American Journal of Roentgenology*, 195(1):260–268, (2010).
- [100] L. M. Wu, J. N. Hu, H. Y. Gu, et al. Can diffusion-weighted MR imaging and contrast-enhanced MR imaging precisely evaluate and predict pathological response to neoadjuvant chemotherapy in patients with breast cancer? *Breast Cancer Research and Treatment*, 135(1):17–28, (2012).
- [101] M. L. Marinovich, N. Houssami, P. MacAskill, et al. Meta-analysis of magnetic resonance imaging in detecting residual breast cancer after neoadjuvant therapy. *Journal of the National Cancer Institute*, 105(5):321–333, (2013).
- [102] S. Sheikhabaei, T. J. Trahan, J. Xiao, et al. FDG-PET/CT and MRI for Evaluation of Pathologic Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer: A Meta-Analysis of Diagnostic Accuracy Studies. *The Oncologist*, 21(8):931–939, (2016).

- [103] R. Croshaw, H. Shapiro-Wright, E. Svensson, K. Erb, and T. Julian. Accuracy of clinical examination, digital mammogram, ultrasound, and MRI in determining postneoadjuvant pathologic tumor response in operable breast cancer patients. In *Annals of Surgical Oncology*, volume 18, pages 3160–3163, (2011).
- [104] P. Therasse, S. G. Arbuck, E. A. Eisenhauer, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *Journal of the National Cancer Institute*, 92(3):205–16, (2000).
- [105] J. R. Scheel, E. Kim, S. C. Partridge, et al. MRI, clinical examination, and mammography for preoperative assessment of residual disease and pathologic complete response after neoadjuvant chemotherapy for breast cancer: ACRIN 6657 trial. *American Journal of Roentgenology*, 210(6):1376–1385, (2018).
- [106] S. Kennedy, M. J. Merino, S. M. Swain, and M. E. Lippman. The effects of hormonal and chemotherapy on tumoral and nonneoplastic breast tissue. *Human Pathology*, 21(2):192–198, (1990).
- [107] K. Wasser, H. P. Sinn, C. Fink, et al. Accuracy of tumor size measurement in breast cancer using MRI is influenced by histological regression induced by neoadjuvant chemotherapy. *European Radiology*, 13(6):1213–1223, (2003).
- [108] F. Denis, A. V. Desbiez-Bourcier, C. Chapiron, et al. Contrast enhanced magnetic resonance imaging underestimates residual disease following neoadjuvant docetaxel based chemotherapy for breast cancer. *European Journal of Surgical Oncology*, 30(10):1069–1076, (2004).
- [109] S. Bahri, J. H. Chen, R. S. Mehta, et al. Residual breast cancer diagnosed by MRI in patients receiving neoadjuvant chemotherapy with and without bevacizumab. *Annals of Surgical Oncology*, 16(6):1619–1628, (2009).
- [110] R. A. Mukhtar, C. Yau, M. Rosen, et al. Clinically Meaningful Tumor Reduction Rates Vary by Prechemotherapy MRI Phenotype and Tumor Subtype in the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Annals of Surgical Oncology*, 20(12):3823–3830, (2013).
- [111] K. P. McGuire, J. Toro-Burguete, H. Dang, et al. MRI staging after neoadjuvant chemotherapy for breast cancer: Does tumor biology affect accuracy? In *Annals of Surgical Oncology*, volume 18, pages 3149–3154, (2011).
- [112] M. L. Marinovich, F. Sardanelli, S. Ciatto, et al. Early prediction of pathologic response to neoadjuvant therapy in breast cancer: Systematic review of the accuracy of MRI. *Breast*, 21(5):669–677, (2012).
- [113] N. M. Hylton, J. D. Blume, W. K. Bernreuter, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy - Results from ACRIN 6657/I-SPY TRIAL. *Radiology*, 263(3):663–672, (2012).

- [114] N. M. Hylton, C. A. Gatsonis, M. A. Rosen, et al. Neoadjuvant chemotherapy for breast cancer: Functional tumor volume by MR imaging predicts recurrence-free survival-results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. *Radiology*, 279(1):44–55, (2016).
- [115] B. Reig, L. Heacock, A. Lewin, N. Cho, and L. Moy. Role of MRI to Assess Response to Neoadjuvant Therapy for Breast Cancer. *Journal of Magnetic Resonance Imaging*, (2020).
- [116] R. Didonato, N. Shapiro, T. Koenigsberg, et al. Invasive mucinous carcinoma of the breast and response patterns after neoadjuvant chemotherapy (NAC). *Histopathology*, 72(6):965–973, (2018).
- [117] D. K. Woolf, A. R. Padhani, N. J. Taylor, et al. Assessing response in breast cancer with dynamic contrast-enhanced magnetic resonance imaging: Are signal intensity–time curves adequate? *Breast Cancer Research and Treatment*, 147(2):335–343, (2014).
- [118] S. Y. Kim, N. Cho, S. U. Shin, et al. Contrast-enhanced MRI after neoadjuvant chemotherapy of breast cancer: lesion-to-background parenchymal signal enhancement ratio for discriminating pathological complete response from minimal residual tumour. *European Radiology*, 28(7):2986–2995, (2018).
- [119] H.-J. Eom, J. H. Cha, W. J. Choi, et al. Predictive Clinicopathologic and Dynamic Contrast-Enhanced MRI Findings for Tumor Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *AJR. American journal of roentgenology*, 208(6):W225–W230, (2017).
- [120] M. L. W. Ah-See, A. Makris, N. J. Taylor, et al. Early changes in functional dynamic magnetic resonance imaging predict for pathologic response to neoadjuvant chemotherapy in primary breast cancer. *Clinical Cancer Research*, 14(20):6580–6589, (2008).
- [121] M. D. Pickles, M. Lowry, D. J. Manton, P. Gibbs, and L. W. Turnbull. Role of dynamic contrast enhanced MRI in monitoring early response of locally advanced breast cancer to neoadjuvant chemotherapy. *Breast Cancer Research and Treatment*, 91(1):1–10, (2005).
- [122] B. E. Dogan, Q. Yuan, R. Bassett, et al. Comparing the Performances of Magnetic Resonance Imaging Size vs Pharmacokinetic Parameters to Predict Response to Neoadjuvant Chemotherapy and Survival in Patients With Breast Cancer. *Current Problems in Diagnostic Radiology*, 48(3):235–240, (2019).
- [123] X. Li, L. R. Arlinghaus, G. D. Ayers, et al. DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: Pilot study findings. *Magnetic Resonance in Medicine*, 71(4):1592–1602, (2014).
- [124] R. Prevos, M. L. Smidt, V. C. Tjan-Heijnen, et al. Pre-treatment differences and early response monitoring of neoadjuvant chemotherapy in breast cancer patients using magnetic resonance imaging: A systematic review. *European Radiology*, 22(12):2607–2616, (2012).

- [125] U. M. Moll and J. Chumas. Morphologic effects of neoadjuvant chemotherapy in locally advanced breast cancer. *Pathology Research and Practice*, 193(3):187–196, (1997).
- [126] R. Rajan, A. Poniecka, T. L. Smith, et al. Change in Tumor Cellularity of Breast Carcinoma after Neoadjuvant Chemotherapy As a Variable in the Pathologic Assessment of Response. *Cancer*, 100(7):1365–1373, (2004).
- [127] M. Briffod, F. Spyrtos, M. Tubiana-Hulin, et al. Sequential cytopunctures during preoperative chemotherapy for primary breast carcinoma. Cytomorphologic changes, initial tumor ploidy, and tumor regression. *Cancer*, 63(4):631–637, (1989).
- [128] S. Kumar, B. A. Badhe, K. Krishnan, and H. Sagili. Study of Tumour Cellularity in Locally Advanced Breast Carcinoma on Neo-Adjuvant Chemotherapy. *Journal of Clinical and Diagnostic Research*, 8(4):FC09, (2014).
- [129] L. Yuan, J. J. Li, C. Q. Li, et al. Diffusion-weighted MR imaging of locally advanced breast carcinoma: The optimal time window of predicting the early response to neoadjuvant chemotherapy. *Cancer Imaging*, 18(1), (2018).
- [130] N. P. Pereira, C. Curi, C. A. Osório, et al. Diffusion-Weighted Magnetic Resonance Imaging of Patients with Breast Cancer Following Neoadjuvant Chemotherapy Provides Early Prediction of Pathological Response – A Prospective Study. *Scientific Reports*, 9(1), (2019).
- [131] S. C. Partridge, Z. Zhang, D. C. Newitt, et al. Diffusion-weighted MRI findings predict pathologic response in neoadjuvant treatment of breast cancer: The ACRIN 6698 multicenter trial. *Radiology*, 289(3):618–627, (2018).
- [132] S. H. Park, W. K. Moon, N. Cho, et al. Comparison of diffusion-weighted MR imaging and FDG PET/CT to predict pathological complete response to neoadjuvant chemotherapy in patients with breast cancer. *European Radiology*, 22(1):18–25, (2012).
- [133] M. D. Pickles, P. Gibbs, M. Lowry, and L. W. Turnbull. Diffusion changes precede size reduction in neoadjuvant treatment of breast cancer. *Magnetic Resonance Imaging*, 24(7):843–847, (2006).
- [134] U. Sharma, K. K. A. Danishad, V. Seenu, and N. R. Jagannathan. Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. *NMR in Biomedicine*, 22(1):104–113, (2009).
- [135] S. H. Park, W. K. Moon, N. Cho, et al. Diffusion-weighted MR imaging: Pretreatment prediction of response to neoadjuvant chemotherapy in patients with breast cancer. *Radiology*, 257(1):56–63, (2010).
- [136] C. Iacconi, M. Giannelli, C. Marini, et al. The role of mean diffusivity (MD) as a predictive index of the response to chemotherapy in locally advanced breast cancer: A preliminary study. *European Radiology*, 20(2):303–308, (2010).

- [137] W. Chu, W. Jin, D. Liu, et al. Diffusion-weighted imaging in identifying breast cancer pathological response to neoadjuvant chemotherapy: A metaanalysis. *Oncotarget*, 9(6):7088–7100, (2018).
- [138] A. B. Ashraf, D. Daye, S. Gavenonis, et al. Identification of intrinsic imaging phenotypes for breast cancer tumors: Preliminary associations with gene expression profiles. *Radiology*, 272(2):374–384, (2014).
- [139] N. M. Braman, M. Etesami, P. Prasanna, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Research*, 19(1):57, (2017).
- [140] F. Aghaei, M. Tan, A. B. Hollingsworth, and B. Zheng. Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy. *Journal of Magnetic Resonance Imaging*, 44(5):1099–1106, (2016).
- [141] E. H. Cain, A. Saha, M. R. Harowicz, et al. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. *Breast cancer research and treatment*, 173(2):455–463, (2019).
- [142] J. Parikh, M. Selmi, G. Charles-Edwards, et al. Changes in Primary Breast Cancer Heterogeneity May Augment Midtreatment MR Imaging Assessment of Response to Neoadjuvant Chemotherapy. *Radiology*, 272(1):100–112, (2014).
- [143] N. L. Eun, D. Kang, E. J. Son, et al. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology*, 294(1):31–41, (2020).
- [144] A. Ahmed, P. Gibbs, M. Pickles, and L. Turnbull. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. *Journal of magnetic resonance imaging*, 38(1):89–101, (2013).
- [145] J. R. Teruel, M. G. Heldahl, P. E. Goa, et al. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *NMR in biomedicine*, 27(8):887–96, (2014).
- [146] F. Chamming's, Y. Ueno, R. Ferré, et al. Features from Computerized Texture Analysis of Breast Cancers at Pretreatment MR Imaging Are Associated with Response to Neoadjuvant Chemotherapy. *Radiology*, 286(2):412–420, (2018).
- [147] S. Chen, Z. Shu, Y. Li, et al. Machine Learning-Based Radiomics Nomogram Using Magnetic Resonance Images for Prediction of Neoadjuvant Chemotherapy Efficacy in Breast Cancer Patients. *Frontiers in Oncology*, 10:1410, (2020).
- [148] R. Ha, C. Chin, J. Karcich, et al. Prior to Initiation of Chemotherapy, Can We Predict Breast Tumor Response? Deep Learning Convolutional Neural Networks Approach Using a Breast MRI Tumor Dataset. *Journal of Digital Imaging*, 32(5):693–701, (2019).

- [149] K. Ravichandran, N. Braman, A. Janowczyk, and A. Madabhushi. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI. page 11. SPIE-Intl Soc Optical Eng, (2018).
- [150] B. Q. Huynh, N. Antropova, and M. L. Giger. Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 101340U. SPIE, (2017).
- [151] Y. H. Qu, H. T. Zhu, K. Cao, et al. Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (DL) method. *Thoracic Cancer*, 11(3):651–658, (2020).
- [152] I. I. Rabi, J. R. Zacharias, S. Millman, and P. Kusch. A new method of measuring nuclear magnetic moment [5]. *Physical Review*, 53(4):318, (1938).
- [153] F. Bloch. Nuclear induction. *Physical Review*, 70(7-8):460–474, (1946).
- [154] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid [7]. *Physical Review*, 69(1-2):37–38, (1946).
- [155] R. Damadian, K. Zaner, D. Hor, and T. DiMaio. Human tumors detected by nuclear magnetic resonance. *Proceedings of the National Academy of Sciences of the United States of America*, 71(4):1471–1473, (1974).
- [156] T. E. Merchant, G. R. P. Thelissen, P. W. de Graaf, et al. Application of a Mixed Imaging Sequence for MR Imaging Characterization of Human Breast Disease. *Acta Radiologica*, 34(4):356–361, (1993).
- [157] R. Rakow-Penner, B. M. Paul Murphy, B. Anders Dale, and B. Haydee Ojeda-Fournier. State of the Art Diffusion Weighted Imaging in the Breast: Recommended Protocol. *Current Radiology Reports*.
- [158] A. N. Nnewihe, T. Grafendorfer, B. L. Daniel, et al. Custom-fitted 16-channel bilateral breast coil for bidirectional parallel imaging. *Magnetic Resonance in Medicine*, 66(1):281–289, (2011).
- [159] Y. Zur, M. L. Wood, and L. J. Neuringer. Spoiling of transverse magnetization in steady-state sequences. *Magnetic Resonance in Medicine*, 21(2):251–263, (1991).
- [160] G. M. Bydder, J. M. Pennock, R. E. Steiner, et al. The short TI inversion recovery sequence-An approach to MR imaging of the abdomen. *Magnetic Resonance Imaging*, 3(3):251–254, (1985).
- [161] J. Frahm, A. Haase, W. Hanicke, et al. Chemical shift selective MR imaging using a whole-body magnet. *Radiology*, 156(2):441–444, (1985).
- [162] T. K. Foo, A. M. Sawyer, W. H. Faulkner, and D. G. Mills. Inversion in the steady state: Contrast optimization and reduced imaging time with fast three-dimensional inversion-recovery-prepared GRE pulse sequences. *Radiology*, 191(1):85–90, (1994).

- [163] M. A. Bernstein, K. F. King, and X. J. Zhou. *Handbook of MRI Pulse Sequences*. Elsevier Inc., (2004).
- [164] W. T. Dixon. Simple proton spectroscopic imaging. *Radiology*, 153(1):189–194, (1984).
- [165] F. Schick, J. Forster, J. Machann, P. Huppert, and C. D. Claussen. Highly selective water and fat imaging applying multislice sequences without sensitivity to B1 field inhomogeneities. *Magnetic Resonance in Medicine*, 38(2):269–274, (1997).
- [166] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: sensitivity encoding for fast MRI. *Magnetic resonance in medicine*, 42(5):952–62, (1999).
- [167] R. M. Mann, C. K. Kuhl, K. Kinkel, and C. Boetes. Breast MRI: Guidelines from the European Society of Breast Imaging. *European Radiology*, 18(7):1307–1318, (2008).
- [168] E. Henderson, B. K. Rutt, and T. Y. Lee. Temporal sampling requirements for the tracer kinetics modeling of breast disease. *Magnetic resonance imaging*, 16(9):1057–73, (1998).
- [169] F. R. Korosec, R. Frayne, T. M. Grist, and C. A. Mistretta. Time-resolved contrast-enhanced 3D MR angiography. *Magnetic resonance in medicine*, 36(3):345–51, (1996).
- [170] L. E. Kershaw and H. L. M. Cheng. A general dual-bolus approach for quantitative DCE-MRI. *Magnetic Resonance Imaging*, 29(2):160–166, (2011).
- [171] E. O. Stejskal and J. E. Tanner. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, 42(1):288–292, (1965).
- [172] J. F. Schenck. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Medical Physics*, 23(6):815–850, (1996).
- [173] D. A. Porter and R. M. Heidemann. High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magnetic Resonance in Medicine*, 62(2):468–475, (2009).
- [174] W. Wu and K. L. Miller. Image formation in diffusion MRI: A review of recent technical developments. *Journal of magnetic resonance imaging : JMRI*, 46(3):646–662, (2017).
- [175] N.-K. Chen, A. Guidon, H.-C. Chang, and A. W. Song. A robust multi-shot scan strategy for high-resolution diffusion weighted MRI enabled by multiplexed sensitivity-encoding (MUSE). *NeuroImage*, 72:41–7, (2013).
- [176] M. Iima, K. Yano, M. Kataoka, et al. Quantitative Non-Gaussian Diffusion and Intravoxel Incoherent Motion Magnetic Resonance Imaging. *Investigative Radiology*, 50(4):205–211, (2015).

- [177] X. Chen, W.-l. Li, Y.-l. Zhang, et al. Meta-analysis of quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesions. *BMC Cancer*, 10(1):693, (2010).
- [178] R.-Y. Shi, Q.-Y. Yao, L.-M. Wu, and J.-R. Xu. Breast Lesions: Diagnosis Using Diffusion Weighted Imaging at 1.5T and 3.0T-Systematic Review and Meta-analysis. *Clinical breast cancer*, 18(3):e305–e320, (2018).
- [179] P. Whiting, A. W. Rutjes, J. B. Reitsma, P. M. Bossuyt, and J. Kleijnen. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, 3(1):25, (2003).
- [180] M. Egger, G. Davey Smith, M. Schneider, and C. Minder. Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical research ed.)*, 315(7109):629–34, (1997).
- [181] J. P. T. Higgins, S. G. Thompson, J. J. Deeks, and D. G. Altman. Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed.)*, 327(7414):557–60, (2003).
- [182] J. B. Reitsma, A. S. Glas, A. W. Rutjes, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10):982–990, (2005).
- [183] E. Eyal, M. Shapiro-Feinberg, E. Furman-Haran, et al. Parametric Diffusion Tensor Imaging of the Breast. *Investigative Radiology*, 47(5):284–291, (2012).
- [184] J. V. Horvat, M. Durando, S. Milans, et al. Apparent diffusion coefficient mapping using diffusion-weighted MRI: impact of background parenchymal enhancement, amount of fibroglandular tissue and menopausal status on breast cancer diagnosis. *European Radiology*, (2018).
- [185] H. Bougias, A. Ghiatas, D. Priovolos, K. Veliou, and A. Christou. Whole-lesion apparent diffusion coefficient (ADC) metrics as a marker of breast tumour characterization—comparison between ADC value and ADC entropy. *The British Journal of Radiology*, 89(1068):20160304, (2016).
- [186] M. Eghtedari, J. Ma, P. Fox, et al. Effects of magnetic field strength and b value on the sensitivity and specificity of quantitative breast diffusion-weighted MRI. *Quantitative Imaging in Medicine and Surgery*, 6(4):374–380, (2016).
- [187] G. Y. Cho, L. Moy, S. G. Kim, et al. Evaluation of breast cancer using intravoxel incoherent motion (IVIM) histogram analysis: comparison with malignant status, histological subtype, and molecular prognostic factors. *European Radiology*, 26(8):2547–2558, (2016).
- [188] H. Dijkstra, M. D. Dorrius, M. Wielema, et al. Quantitative DWI implemented after DCE-MRI yields increased specificity for BI-RADS 3 and 4 breast lesions. *Journal of Magnetic Resonance Imaging*, 44(6):1642–1649, (2016).

- [189] C. Liu, K. Wang, Q. Chan, et al. Intravoxel incoherent motion MR imaging for breast lesions: comparison and correlation with pharmacokinetic evaluation from dynamic contrast-enhanced MR imaging. *European radiology*, 26(11):3888–3898, (2016).
- [190] U. Sharma, R. G. Sah, K. Agarwal, et al. Potential of Diffusion-Weighted Imaging in the Characterization of Malignant, Benign, and Healthy Breast Tissues and Molecular Subtypes of Breast Cancer. *Frontiers in Oncology*, 6:126, (2016).
- [191] O. Arponen, A. Masarwah, A. Sutela, et al. Incidentally detected enhancing lesions found in breast MRI: analysis of apparent diffusion coefficient and T2 signal intensity significantly improves specificity. *European Radiology*, 26(12):4361–4370, (2016).
- [192] L. Si, R. Zhai, X. Liu, et al. MRI in the differential diagnosis of primary architectural distortion detected by mammography. *Diagnostic and Interventional Radiology*, 22(2):141–150, (2016).
- [193] T. Pawlik and J. Ryś. Comparison of apparent diffusion coefficient in diffusion-weighted magnetic resonance imaging and morphological assessment of breast tumors. *Polish Journal of Pathology*, 4(4):398–403, (2016).
- [194] C. S. Guatelli, A. G. V. Bitencourt, C. A. B. d. T. Osório, et al. Can diffusion-weighted imaging add information in the evaluation of breast lesions considered suspicious on magnetic resonance imaging? *Radiologia brasileira*, 50(5):291–298, (2017).
- [195] Y. Akın, M. Uğurlu, H. Kaya, and E. Arıbal. Diagnostic Value of Diffusion-weighted Imaging and Apparent Diffusion Coefficient Values in the Differentiation of Breast Lesions, Histopathologic Subgroups and Correlation with Prognostic Factors using 3.0 Tesla MR. *The journal of breast health*, 12(3):123–132, (2016).
- [196] E. Arıbal, R. Asadov, A. Ramazan, M. Ugurlu, and H. Kaya. Multiparametric breast MRI with 3T: Effectivity of combination of contrast enhanced MRI, DWI and 1H single voxel spectroscopy in differentiation of Breast tumors. *European Journal of Radiology*, 85(5):979–986, (2016).
- [197] A. M. Osman and N. H. Shebrya. Value of diffusion weighted imaging (DWI) and apparent diffusion coefficient factor (ADC) calculation in differentiation of solid breast lesions. *The Egyptian Journal of Radiology and Nuclear Medicine*, 47(1):363–371, (2016).
- [198] R. Fusco, M. Sansone, S. Filice, et al. Integration of DCE-MRI and DW-MRI Quantitative Parameters for Breast Lesion Classification. *BioMed Research International*, 2015:1–12, (2015).
- [199] K. Sun, X. Chen, W. Chai, et al. Breast Cancer: Diffusion Kurtosis MR Imaging—Diagnostic Accuracy and Correlation with Clinical-Pathologic Factors. *Radiology*, 277(1):46–55, (2015).
- [200] Q. Min, K. Shao, L. Zhai, et al. Differential diagnosis of benign and malignant breast masses using diffusion-weighted magnetic resonance imaging. *World journal of surgical oncology*, 13:32, (2015).

- [201] S. Kul, Oğuz, Eyüboğlu, and Kömürçüoğlu. Can unenhanced breast MRI be used to decrease negative biopsy rates? *Diagnostic and interventional radiology (Ankara, Turkey)*, 21(4):287–92, (2015).
- [202] R. Caivano, A. Villonio, F. D’Antuono, et al. Diffusion Weighted Imaging and Apparent Diffusion Coefficient in 3 Tesla Magnetic Resonance Imaging of Breast Lesions. *Cancer Investigation*, 33(5):159–164, (2015).
- [203] G. Çabuk, M. Nass Duce, A. Özgür, et al. The diagnostic value of diffusion-weighted imaging and the apparent diffusion coefficient values in the differentiation of benign and malignant breast lesions. *Journal of Medical Imaging and Radiation Oncology*, 59(2):141–148, (2015).
- [204] Y. A. Ibrahim, L. Habib, and A. Deif. Role of quantitative diffusion weighted imaging in characterization of breast masses. *The Egyptian Journal of Radiology and Nuclear Medicine*, 46(3):805–810, (2015).
- [205] M. H. Abowarda, D. I. Hasan, and O. A. Elteeh. Predictive value of ADC mapping in discriminating probably benign and suspicious breast lesions. *The Egyptian Journal of Radiology and Nuclear Medicine*, 46(2):545–551, (2015).
- [206] M. A. K. A. Wahab, H. A. Kareem, and E. E. Hassan. The utility of diffusion weighted MRI and apparent diffusion coefficient in characterization of breast masses. *The Egyptian Journal of Radiology and Nuclear Medicine*, 46(4):1257–1265, (2015).
- [207] I. M. El Fiki, H. M. Abdel-Rahman, and M. M. Morsy. Assessment of breast mass: Utility of diffusion-weighted MR and MR spectroscopy imaging. *The Egyptian Journal of Radiology and Nuclear Medicine*, 46(4):1327–1335, (2015).
- [208] C. Altay, P. Balci, S. Altay, et al. Diffusion-weighted MR imaging: role in the differential diagnosis of breast lesions. *JBR-BTR : organe de la Societe royale belge de radiologie (SRBR) = orgaan van de Koninklijke Belgische Vereniging voor Radiologie (KBVR)*, 97(4):211–6, (2014).
- [209] F. Z. Moukhtar and A. A. Abu El Maati. Apparent diffusion coefficient values as an adjunct to dynamic contrast enhanced MRI for discriminating benign and malignant breast lesions presenting as mass and non-mass like enhancement. *The Egyptian Journal of Radiology and Nuclear Medicine*, 45(2):597–604, (2014).
- [210] S. Kul, I. Eyuboglu, A. Cansu, and E. Alhan. Diagnostic efficacy of the diffusion weighted imaging in the characterization of different types of breast lesions. *Journal of Magnetic Resonance Imaging*, 40(5):1158–1164, (2014).
- [211] H. Yoo, H. J. Shin, S. Baek, et al. Diagnostic performance of apparent diffusion coefficient and quantitative kinetic parameters for predicting additional malignancy in patients with newly diagnosed breast cancer. *Magnetic Resonance Imaging*, 32(7):867–874, (2014).
- [212] H. Cai, L. Liu, Y. Peng, Y. Wu, and L. Li. Diagnostic assessment by dynamic contrast-enhanced and diffusion-weighted magnetic resonance in differentiation of breast lesions under different imaging protocols. *BMC Cancer*, 14(1):366, (2014).

- [213] C. Spick, K. Pinker-Domenig, M. Rudas, T. H. Helbich, and P. A. Baltzer. MRI-only lesions: application of diffusion-weighted imaging obviates unnecessary MR-guided breast biopsies. *European Radiology*, 24(6):1204–1210, (2014).
- [214] S. Kothari, A. Singh, U. Das, et al. Role of exponential apparent diffusion coefficient in characterizing breast lesions by 3.0 Tesla diffusion-weighted magnetic resonance imaging. *Indian Journal of Radiology and Imaging*, 27(2):229, (2017).
- [215] L. Nogueira, S. Brandão, E. Matos, et al. Diffusion-weighted breast imaging at 3 T: Preliminary experience. *Clinical Radiology*, 69(4):378–384, (2014).
- [216] L. Bokacheva, J. B. Kaplan, D. D. Giri, et al. Intravoxel incoherent motion diffusion-weighted MRI at 3.0 T differentiates malignant breast lesions from benign lesions and breast parenchyma. *Journal of Magnetic Resonance Imaging*, 40(4):813–823, (2014).
- [217] S. Tan, K. Rahmat, F. Rozalli, et al. Differentiation between benign and malignant breast lesions using quantitative diffusion-weighted sequence on 3 T MRI. *Clinical Radiology*, 69(1):63–71, (2014).
- [218] H. H. M. Hassan, M. H. Mahmoud Zahran, H. El-Prince Hassan, A.-E. Mohamed Abdel-Hamid, and G. Abdel Shafy Fadaly. Diffusion magnetic resonance imaging of breast lesions: Initial experience at Alexandria University. *Alexandria Journal of Medicine*, 49(3):265–272, (2013).
- [219] C. Liu, C. Liang, Z. Liu, S. Zhang, and B. Huang. Intravoxel incoherent motion (IVIM) in evaluation of breast lesions: Comparison with conventional DWI. *European Journal of Radiology*, 82(12):e782–e789, (2013).
- [220] L. Cheng, Y. Bai, J. Zhang, et al. Optimization of apparent diffusion coefficient measured by diffusion-weighted MRI for diagnosis of breast lesions presenting as mass and non-mass-like enhancement. *Tumor Biology*, 34(3):1537–1545, (2013).
- [221] M. Ochi, T. Kuroiwa, S. Sunami, et al. Diffusion-weighted imaging (b value = 1500 s/mm²) is useful to decrease false-positive breast cancer cases due to fibrocystic changes. *Breast Cancer*, 20(2):137–144, (2013).
- [222] S. Orguc, I. Basara, and T. Coskun. Diffusion-weighted MR imaging of the breast: comparison of apparent diffusion coefficient values of normal breast tissue with benign and malignant breast lesions. *Singapore medical journal*, 53(11):737–43, (2012).
- [223] X. Chen, X.-J. He, R. Jin, et al. Conspicuity of breast lesions at different b values on diffusion-weighted imaging. *BMC Cancer*, 12(1):334, (2012).
- [224] J. R. M. de Almeida, A. B. Gomes, T. P. Barros, P. E. Fahel, and M. d. S. Rocha. Diffusion-weighted imaging of suspicious (BI-RADS 4) breast lesions: stratification based on histopathology. *Radiologia brasileira*, 50(3):154–161, (2017).
- [225] B. Zhang, B. Zhu, M. Li, et al. Comparative utility of MRI perfusion with MSIDR and DWIBS for the characterization of breast tumors. *Acta Radiologica*, 53(6):607–614, (2012).

- [226] G. Sonmez, F. Cuce, H. Mutlu, et al. Value of diffusion-weighted MRI in the differentiation of benign and malign breast lesions. *Wiener klinische Wochenschrift*, 123(21-22):655–661, (2011).
- [227] K. Inoue, E. Kozawa, W. Mizukoshi, et al. Usefulness of diffusion-weighted imaging of breast tumors: quantitative and visual assessment. *Japanese Journal of Radiology*, 29(6):429–436, (2011).
- [228] S. Kul, A. Cansu, E. Alhan, et al. Contribution of Diffusion-Weighted Imaging to Dynamic Contrast-Enhanced MRI in the Characterization of Breast Tumors. *American Journal of Roentgenology*, 196(1):210–217, (2011).
- [229] F. Fornasa, L. Pinali, A. Gasparini, E. Toniolli, and S. Montemezzi. Diffusion-weighted magnetic resonance imaging in focal breast lesions: analysis of 78 cases with pathological correlation. *La radiologia medica*, 116(2):264–275, (2011).
- [230] G. K. Gouhar and E.-S. H. Zidan. Diffusion-weighted imaging of breast tumors: Differentiation of benign and malignant tumors. *The Egyptian Journal of Radiology and Nuclear Medicine*, 42(1):93–99, (2011).
- [231] T. Imamura, I. Isomoto, E. Sueyoshi, et al. Diagnostic performance of ADC for Non-mass-like breast lesions on MR imaging. *Magnetic resonance in medical sciences : MRMS : an official journal of Japan Society of Magnetic Resonance in Medicine*, 9(4):217–25, (2010).
- [232] G. Jin, N. An, M. A. Jacobs, and K. Li. The role of parallel diffusion-weighted imaging and apparent diffusion coefficient (ADC) map values for evaluating breast lesions: preliminary results. *Academic radiology*, 17(4):456–63, (2010).
- [233] S. C. Partridge, W. B. DeMartini, B. F. Kurland, et al. Differential diagnosis of mammographically and clinically occult breast lesions on diffusion-weighted MRI. *Journal of Magnetic Resonance Imaging*, 31(3):562–570, (2010).
- [234] P. Belli, M. Costantini, E. Bufi, et al. Diffusion-weighted imaging in breast lesion evaluation. *La radiologia medica*, 115(1):51–69, (2010).
- [235] H. Yabuuchi, Y. Matsuo, T. Kamitani, et al. Non-mass-like enhancement on contrast-enhanced breast MR imaging: Lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. *European Journal of Radiology*, 75(1):e126–e132, (2010).
- [236] L. Palle and B. Reddy. Role of diffusion MRI in characterizing benign and malignant breast lesions. *The Indian journal of radiology & imaging*, 19(4):287–90, (2009).
- [237] F. P. A. Pereira, G. Martins, E. Figueiredo, et al. Assessment of Breast Lesions With Diffusion-Weighted MRI: Comparing the Use of Different b Values. *American Journal of Roentgenology*, 193(4):1030–1035, (2009).
- [238] M. Tozaki and E. Fukuma. ¹H MR Spectroscopy and Diffusion-Weighted Imaging of the Breast: Are They Useful Tools for Characterizing Breast Lesions Before Biopsy? *American Journal of Roentgenology*, 193(3):840–849, (2009).

- [239] G. G. Lo, V. Ai, J. K. Chan, et al. Diffusion-Weighted Magnetic Resonance Imaging of Breast Lesions. *Journal of Computer Assisted Tomography*, 33(1):63–69, (2009).
- [240] H. Yabuuchi, Y. Matsuo, T. Okafuji, et al. Enhanced mass on contrast-enhanced breast MR imaging: Lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. *Journal of Magnetic Resonance Imaging*, 28(5):1157–1165, (2008).
- [241] E. Wenkel, C. Geppert, R. Schulz-Wendtland, et al. Diffusion Weighted Imaging in Breast MRI. *Academic Radiology*, 14(9):1077–1083, (2007).
- [242] E. Rubesova, A.-S. Grell, V. De Maertelaer, et al. Quantitative diffusion imaging in breast cancer: A clinical prospective study. *Journal of Magnetic Resonance Imaging*, 24(2):319–324, (2006).
- [243] R. Woodhams, K. Matsunaga, S. Kan, et al. ADC mapping of benign and malignant breast tumors. *Magnetic resonance in medical sciences : MRMS : an official journal of Japan Society of Magnetic Resonance in Medicine*, 4(1):35–42, (2005).
- [244] N. Lin, J. Chen, J. Hua, et al. Intravoxel incoherent motion MR imaging in breast cancer: quantitative analysis for characterizing lesions. *Int J Clin Exp Med*, 10(1):1705–1714, (2017).
- [245] Q. Wang, Y. Guo, J. Zhang, et al. Contribution of IVIM to Conventional Dynamic Contrast-Enhanced and Diffusion-Weighted MRI in Differentiating Benign from Malignant Breast Masses. *Breast care (Basel, Switzerland)*, 11(4):254–258, (2016).
- [246] L. Jiang, X. Lu, B. Hua, et al. Intravoxel Incoherent Motion Diffusion-Weighted Imaging Versus Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Journal of Computer Assisted Tomography*, 42(1):1, (2017).
- [247] M. Iima, M. Kataoka, S. Kanao, et al. Intravoxel Incoherent Motion and Quantitative Non-Gaussian Diffusion MR Imaging: Evaluation of the Diagnostic and Prognostic Value of Several Markers of Malignant and Benign Breast Lesions. *Radiology*, page 162853, (2017).
- [248] N. H. Peters, K. L. Vincken, M. A. van den Bosch, et al. Quantitative diffusion weighted imaging for differentiation of benign and malignant breast lesions: The influence of the choice of b-values. *Journal of Magnetic Resonance Imaging*, 31(5):1100–1105, (2010).
- [249] W. Bogner, S. Gruber, K. Pinker, et al. Diffusion-weighted MR for Differentiation of Breast Lesions at 3.0 T: How Does Selection of Diffusion Protocols Affect Diagnosis? *Radiology*, 253(2):341–351, (2009).
- [250] M. D. Dorrius, H. Dijkstra, M. Oudkerk, and P. E. Sijens. Effect of b value and pre-admission of contrast on diagnostic accuracy of 1.5-T breast DWI: a systematic review and meta-analysis. *European Radiology*, 24(11):2835–2847, (2014).
- [251] C. Federau. Intravoxel incoherent motion MRI as a means to measure *in vivo* perfusion: A review of the evidence. *NMR in Biomedicine*, 30(11):e3780, (2017).

- [252] S. Brandão, L. Nogueira, E. Matos, et al. Fat suppression techniques (STIR vs. SPAIR) on diffusion-weighted imaging of breast lesions at 3.0 T: preliminary experience. *La radiologia medica*, 120(8):705–713, (2015).
- [253] G. C. Baxter, A. J. Patterson, R. Woitek, et al. Improving the image quality of DWI in breast cancer: comparison of multi-shot DWI using multiplexed sensitivity encoding to conventional single-shot echo-planar imaging DWI. *The British Journal of Radiology*, page 20200427, (2020).
- [254] D. J. Wisner, N. Rogers, V. S. Deshpande, et al. High-resolution diffusion-weighted imaging for the separation of benign from malignant BI-RADS 4/5 lesions found on breast MRI at 3T. *Journal of Magnetic Resonance Imaging*, 40(3):674–681, (2014).
- [255] Y. J. Kim, S. H. Kim, B. J. Kang, et al. Readout-segmented echo-planar imaging in diffusion-weighted mr imaging in breast cancer: comparison with single-shot echo-planar imaging in image quality. *Korean journal of radiology*, 15(4):403–10, (2014).
- [256] Y. Y. An, S. H. Kim, and B. J. Kang. Differentiation of malignant and benign breast lesions: Added value of the qualitative analysis of breast lesions on diffusion-weighted imaging (DWI) using readout-segmented echo-planar imaging at 3.0 T. *PLOS ONE*, 12(3):e0174681, (2017).
- [257] W. Bogner, K. Pinker-Domenig, H. Bickel, et al. Readout-segmented echo-planar imaging improves the diagnostic performance of diffusion-weighted MR breast examinations at 3.0 T. *Radiology*, 263(1):64–76, (2012).
- [258] F. Crété-Roffet, T. Dolmiere, P. Ladret, M. Nicolas, and F. Crete. The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric. SPIE Electronic Imaging Symposium Conf Human Vision and Electronic Imagin. Technical report, (2007).
- [259] J. A. McKay, G. J. Metzger, S. Moeller, et al. Adaptation of a Computer Vision Blur Metric to Objectively Compare High Resolution DWI Strategies in in vivo Breast Imaging. In *Proc. Intl. Soc. Mag. Reson. Med.* 27, (2019).
- [260] D. Mattes, D. Haynor, H. Vesselle, T. Lewellen, and W. Eubank. PET-CT image registration in the chest using free-form deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128, (2003).
- [261] K. E. Keenan, L. J. Wilmes, S. O. Aliu, et al. Design of a breast phantom for quantitative MRI. *Journal of Magnetic Resonance Imaging*, 44(3):610–619, (2016).
- [262] H. C. Chang, P. Gaur, Y. H. Chou, M. L. Chu, and N. K. Chen. Interleaved EPI based fMRI improved by multiplexed sensitivity encoding (MUSE) and simultaneous multi-band imaging. *PLoS ONE*, 9(12), (2014).
- [263] H. C. Chang, M. Sundman, L. Petit, et al. Human brain diffusion tensor imaging at submillimeter isotropic resolution on a 3Tesla clinical MRI scanner. *NeuroImage*, 118:667–675, (2015).

- [264] J. R. Teruel, H. E. Fjøsne, A. Østlie, et al. Inhomogeneous static magnetic field-induced distortion correction applied to diffusion weighted MRI of the breast at 3T. *Magnetic Resonance in Medicine*, 74(4):1138–1144, (2015).
- [265] L. Filli, S. Ghafoor, D. Kenkel, et al. Simultaneous multi-slice readout-segmented echo planar imaging for accelerated diffusion-weighted imaging of the breast. *European Journal of Radiology*, 85(1):274–278, (2016).
- [266] R. M. Mann, C. Balleyguier, P. A. Baltzer, et al. Breast MRI: EUSOBI recommendations for women’s information. *European Radiology*, 25(12):3669–3678, (2015).
- [267] G. Baxter, A. Selamoglu, J. Mackay, et al. A meta-analysis comparing the diagnostic performance of abbreviated MRI and a full diagnostic protocol in breast cancer. *Clinical Radiology*, (2020).
- [268] V. L. Mango, E. A. Morris, D. David Dershaw, et al. Abbreviated protocol for breast MRI: Are multiple sequences needed for cancer detection? *European Journal of Radiology*, 84(1):65–70, (2015).
- [269] L. Heacock, A. N. Melsaether, S. L. Heller, et al. Evaluation of a known breast cancer using an abbreviated breast MRI protocol: Correlation of imaging characteristics and pathology with lesion detection and conspicuity. *European Journal of Radiology*, 85(4):815–823, (2016).
- [270] M. Moschetta, M. Telegrafo, L. Rella, A. A. Stabile Ianora, and G. Angelelli. Abbreviated Combined MR Protocol: A New Faster Strategy for Characterizing Breast Lesions. *Clinical Breast Cancer*, 16(3):207–211, (2016).
- [271] G. Oldrini, B. Fedida, J. Poujol, et al. Abbreviated breast magnetic resonance protocol: Value of high-resolution temporal dynamic sequence to improve lesion characterization. *European Journal of Radiology*, 95:177–185, (2017).
- [272] E. S. Ko and E. A. Morris. Abbreviated magnetic resonance imaging for breast cancer screening: Concept, early results, and considerations. *Korean Journal of Radiology*, 20(4):533–541, (2019).
- [273] A. R. Mootz, A. J. Madhuranthakam, and B. Dogan. Changing Paradigms in Breast Cancer Screening: Abbreviated Breast MRI. *European Journal of Breast Health*, 15(1):1–6, (2019).
- [274] H. I. Greenwood. Abbreviated protocol breast MRI: The past, present, and future. *Clinical Imaging*, 53:169–173, (2019).
- [275] K. Deike-Hofmann, F. Koenig, D. Paech, et al. Abbreviated MRI Protocols in Breast Cancer Diagnostics. *Journal of Magnetic Resonance Imaging*, 49(3):647–658, (2019).
- [276] S. C. Harvey, P. A. Di Carlo, B. Lee, et al. An Abbreviated Protocol for High-Risk Screening Breast MRI Saves Time and Resources. *Journal of the American College of Radiology*, 13(11):R74–R80, (2016).
- [277] D. Leithner, G. J. Wengert, T. H. Helbich, et al. Clinical role of breast MRI now and going forward. *Clinical Radiology*, 73(8):700–714, (2018).

- [278] C. M. Chhor and C. L. Mercado. Abbreviated MRI protocols: Wave of the future for breast cancer screening. *American Journal of Roentgenology*, 208(2):284–289, (2017).
- [279] M. D. McInnes, D. Moher, B. D. Thombs, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies The PRISMA-DTA Statement. *JAMA - Journal of the American Medical Association*, 319(4):388–396, (2018).
- [280] S.-Q. Chen, M. Huang, Y.-Y. Shen, C.-L. Liu, and C.-X. Xu. Abbreviated MRI Protocols for Detecting Breast Cancer in Women with Dense Breasts. *Korean journal of radiology*, 18(3):470–475, (2017).
- [281] A. Petrillo, R. Fusco, M. Sansone, et al. Abbreviated breast dynamic contrast-enhanced MR imaging for lesion detection and characterization: the experience of an Italian oncologic center. *Breast Cancer Research and Treatment*, 164(2):401–410, (2017).
- [282] Y. Machida, A. Shimauchi, Y. Kanemaki, et al. Feasibility and potential limitations of abbreviated breast MRI: an observer study using an enriched cohort. *Breast Cancer*, 24(3):411–419, (2017).
- [283] B. Panigrahi, L. Mullen, E. Falomo, B. Panigrahi, and S. Harvey. An Abbreviated Protocol for High-risk Screening Breast Magnetic Resonance Imaging. *Academic Radiology*, 24(9):1132–1138, (2017).
- [284] N. Seppala, R. Fallah Rastegar, L. Richmond, et al. Rapid MRI of the breast in evaluating lesions discovered on screening. *Breast Journal*, 24(6):986–991, (2018).
- [285] V. Romeo, R. Cuocolo, R. Liuzzi, et al. Preliminary Results of a Simplified Breast MRI Protocol to Characterize Breast Lesions. *Academic Radiology*, 24(11):1387–1394, (2017).
- [286] S. Bickelhaupt, J. Tesdorff, F. B. Laun, et al. Independent value of image fusion in unenhanced breast MRI using diffusion-weighted and morphological T2-weighted images for lesion characterization in patients with recently detected BI-RADS 4/5 x-ray mammography findings. *European Radiology*, 27(2):562–569, (2017).
- [287] P. T. Von Hippel. The heterogeneity statistic I² can be biased in small meta-analyses. *BMC Medical Research Methodology*, 15(1), (2015).
- [288] C. K. Kuhl. Abbreviated breast MRI for screening women with dense breast: The EA1141 trial. *British Journal of Radiology*, 91(1090), (2018).
- [289] Y. Takwoingi, M. M. Leeftang, and J. J. Deeks. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Annals of Internal Medicine*, 158(7):544–554, (2013).
- [290] A. Bardia and J. Baselga. Neoadjuvant therapy as a platform for drug development and approval in breast cancer. *Clinical Cancer Research*, 19(23):6360–6370, (2013).

- [291] L. J. Esserman, D. A. Berry, A. DeMichele, et al. Pathologic complete response predicts recurrence-free survival more effectively by cancer subset: results from the I-SPY 1 TRIAL–CALGB 150007/150012, ACRIN 6657. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 30(26):3242–9, (2012).
- [292] D. Yee, A. M. DeMichele, C. Yau, et al. Association of Event-Free and Distant Recurrence–Free Survival With Individual-Level Pathologic Complete Response in Neoadjuvant Treatment of Stages 2 and 3 Breast Cancer. *JAMA Oncology*, 6(9):1355, (2020).
- [293] Y. L. Gu, S. M. Pan, J. Ren, Z. X. Yang, and G. Q. Jiang. Role of Magnetic Resonance Imaging in Detection of Pathologic Complete Remission in Breast Cancer Patients Treated With Neoadjuvant Chemotherapy: A Meta-analysis. *Clinical Breast Cancer*, 17(4):245–255, (2017).
- [294] A. A. Onitilo, J. M. Engel, R. T. Greenlee, and B. N. Mukesh. Breast Cancer Subtypes Based on ER/PR and Her2 Expression: Comparison of Clinicopathologic Features and Survival. *Clinical Medicine & Research*, 7(1-2):4–13, (2009).
- [295] S. E. Pinder, E. Provenzano, H. Earl, and I. O. Ellis. Laboratory handling and histology reporting of breast specimens from patients who have received neoadjuvant chemotherapy. *Histopathology*, 50(4):409–417, (2007).
- [296] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek. The Design of SimpleITK. *Frontiers in Neuroinformatics*, 7(DEC):45, (2013).
- [297] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare. SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *Journal of Digital Imaging*, 31(3):290–303, (2018).
- [298] N. J. Tustison, B. B. Avants, P. A. Cook, et al. N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, (2010).
- [299] J. J. Van Griethuysen, A. Fedorov, C. Parmar, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21):e104–e107, (2017).
- [300] R. Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, (1996).
- [301] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, (1995).
- [302] T. E. Oliphant. Python for scientific computing. *Computing in Science and Engineering*, 9(3):10–20, (2007).
- [303] I. Banerjee, S. Malladi, D. Lee, et al. Assessing treatment response in triple-negative breast cancer from quantitative image analysis in perfusion magnetic resonance imaging. *Journal of Medical Imaging*, 5(01):1, (2017).

- [304] D. I. Golden, J. A. Lipson, M. L. Telli, J. M. Ford, and D. L. Rubin. Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer. *Journal of the American Medical Informatics Association*, 20(6):1059–1066, (2013).
- [305] W. J. Fu. Penalized regressions: The bridge versus the lasso? *Journal of Computational and Graphical Statistics*, 7(3):397–416, (1998).
- [306] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2):301–320, (2005).
- [307] F. E. Harrell, K. L. Lee, and D. B. Mark. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4):361–387, (1996).
- [308] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12):1373–1379, (1996).
- [309] W. Li, D. C. Newitt, B. L. Yun, et al. Tumor sphericity predicts response in neoadjuvant chemotherapy for invasive breast cancer. *Tomography*, 6(2):216–222, (2020).
- [310] Y. C. Chang, C. S. Huang, Y. J. Liu, et al. Angiogenic response of locally advanced breast cancer to neoadjuvant chemotherapy evaluated with parametric histogram from dynamic contrast-enhanced MRI. *Physics in Medicine and Biology*, 49(16):3593–3602, (2004).
- [311] M. M. Nadrljanski and Z. C. Milosevic. Tumor texture parameters of invasive ductal breast carcinoma in neoadjuvant chemotherapy: early identification of non-responders on breast MRI. *Clinical Imaging*, 65:119–123, (2020).
- [312] J. Lee, S. H. Kim, and B. J. Kang. Pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: Perfusion metrics of dynamic contrast enhanced MRI. *Scientific reports*, 8(1):9490, (2018).
- [313] Z. Liu, Z. Li, J. Qu, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: A multicenter study. *Clinical Cancer Research*, 25(12):3538–3547, (2019).
- [314] C. E. Loo, H. J. Teertstra, S. Rodenhuis, et al. Dynamic contrast-enhanced MRI for prediction of breast cancer response to neoadjuvant chemotherapy: Initial results. *American Journal of Roentgenology*, 191(5):1331–1338, (2008).
- [315] A. Rieber, H. Zeitler, H. Rosenthal, et al. MRI of breast cancer: Influence of chemotherapy on sensitivity. *British Journal of Radiology*, 70(MAY):452–458, (1997).
- [316] C. El Khoury, V. Servois, F. Thibault, et al. MR Quantification of the Washout Changes in Breast Tumors Under Preoperative Chemotherapy: Feasibility and Preliminary Results. *American Journal of Roentgenology*, 184(5):1499–1504, (2005).

- [317] X. Li, L. R. Arlinghaus, G. D. Ayers, et al. DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: Pilot study findings. *Magnetic Resonance in Medicine*, 71(4):1592–1602, (2014).
- [318] M. M. Panzeri, C. Losio, A. Della Corte, et al. Prediction of Chemoresistance in Women Undergoing Neo-Adjuvant Chemotherapy for Locally Advanced Breast Cancer: Volumetric Analysis of First-Order Textural Features Extracted from Multiparametric MRI. *Contrast Media and Molecular Imaging*, 2018, (2018).
- [319] R. W. Granzier, N. M. Verbakel, A. Ibrahim, et al. MRI-based radiomics in breast cancer: feature robustness with respect to inter-observer segmentation variability. *Scientific Reports*, 10(1):14163, (2020).
- [320] E. J. Sutton, N. Onishi, D. A. Fehr, et al. A machine learning model that classifies breast cancer pathologic complete response on MRI post-neoadjuvant chemotherapy. *Breast Cancer Research*, 22(1):57, (2020).
- [321] A. Tahmassebi, G. J. Wengert, T. H. Helbich, et al. Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients. *Investigative Radiology*, 54(2):110–117, (2019).
- [322] P. Baltzer, R. M. Mann, M. Iima, et al. Diffusion-weighted imaging of the breast-a consensus and mission statement from the EUSOBI International Breast Diffusion-Weighted Imaging working group On behalf of the EUSOBI international Breast Diffusion-Weighted Imaging working group. *European Radiology*, 30:1436–1450, (2020).
- [323] T. Li, Y. Hong, D. Kong, and K. Li. Histogram analysis of diffusion kurtosis imaging based on whole-volume images of breast lesions. *Journal of Magnetic Resonance Imaging*, 51(2):627–634, (2020).
- [324] W. Liu, C. Wei, J. Bai, X. Gao, and L. Zhou. Histogram analysis of diffusion kurtosis imaging in the differentiation of malignant from benign breast lesions. *European Journal of Radiology*, 117:156–163, (2019).
- [325] N. Meng, X. Wang, J. Sun, et al. A comparative study of the value of amide proton transfer-weighted imaging and diffusion kurtosis imaging in the diagnosis and evaluation of breast cancer. *European Radiology*, (2020).
- [326] J. A. McKay, A. L. Church, N. Rubin, et al. A Comparison of Methods for High-Spatial-Resolution Diffusion-weighted Imaging in Breast MRI. *Radiology*, page 200221, (2020).
- [327] M. W. Barentsz, V. Taviani, J. M. Chang, et al. Assessment of tumor morphology on diffusion-weighted (DWI) breast MRI: Diagnostic value of reduced field of view DWI. *Journal of Magnetic Resonance Imaging*, 42(6):1656–1665, (2015).
- [328] J. Y. Park, H. J. Shin, K. C. Shin, et al. Comparison of readout segmented echo planar imaging (EPI) and EPI with reduced field-of-view diffusion-weighted imaging at 3t in patients with breast cancer. *Journal of Magnetic Resonance Imaging*, 42(6):1679–1688, (2015).

- [329] K. L. Granlund, E. Staroswiecki, M. T. Alley, B. L. Daniel, and B. A. Hargreaves. High-resolution, three-dimensional diffusion-weighted breast imaging using DESS. *Magnetic Resonance Imaging*, 32(4):330–341, (2014).
- [330] ClinicalTrials.gov. Breast Screening - Risk Adaptive Imaging for Density (BRAID).
- [331] F. Pedregosa, V. Michel, O. Grisel, et al. Scikit-learn: Machine Learning in Python. Technical report, (2011).
- [332] A. Paszke, S. Gross, F. Massa, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Technical report, (2019).
- [333] M. Abadi, A. Agarwal, P. Barham, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. (2016).
- [334] D. Newitt and N. Hylton. Multi-center breast DCE-MRI data and segmentations from patients in the I-SPY 1/ACRIN 6657 trials.
- [335] M. E. Mayerhoefer, P. Szomolanyi, D. Jirak, A. Materka, and S. Trattnig. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: An application-oriented study. *Medical Physics*, 36(4):1236–1243, (2009).
- [336] S. Rizzo, F. Botta, S. Raimondi, et al. Radiomics: the facts and the challenges of image analysis. *European Radiology Experimental*, 2(1), (2018).
- [337] A. Zwanenburg, M. Vallières, M. A. Abdalah, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338, (2020).

Appendix 1

First Author	Diffusion Model	Journal	n	MR vendor	Field strength (T)	Breast coil	b-values (s/mm ²)	TR/TE	Matrix size	Slice thickness (mm)	FOV (mm ²)
Horvat 2018 [184]	ADC	European Radiology	288	GE	3	16-channel	0, 600	6000/minimum	98x98 or 128x128	4 or 5	(280-380) x (280-380)
Guatelli 2017 [194]	ADC	Radiologia Brasileira	215	GE	1.5		0, 750	4000/94	192x192	3	
Yamaguchi 2017 [58]	ADC, DTI	Magnetic Resonance Medical Sciences	80	Siemens	1.5	2-channel	0, 1000	5500/75 (ss-EPI), 3000/69 (DTI)	140x140	5	350x350
Kothari 2017 [214]	ADC	Indian Journal of Radiological Imaging	77	GE	3	16-channel	0, 800	5825/minimum	96x140	4	320x320
de Almeida 2017 [224]	ADC	Radiologia Brasileira	76	GE	1.5	8-channel	0, 750	11700/96	256x224	3.5	340x340
Suo 2017 [56]	ADC, IVIM	Journal of Magnetic Resonance Imaging	101	Philips	3		0, 10, 30, 50, 100, 150, 200, 500, 800, 1000, 1500, 2000, 2500	4500/85	108x128	3	280x330
Onaygil 2017 [62]	DTI	Journal of Magnetic Resonance Imaging	85	Siemens	3	16-channel	50, 400, 800	9700/86 (DWI), 6400/76 (DTI)	192x82 (DWI), 164x96 (DTI)	3	320x165 (DWI), 320x180 (DTI)
Lin 2017 [244]	ADC	International Journal of Clinical and Experimental Medicine	93	Philips	3	4-channel	0, 50, 100, 150, 200, 500, 800	2000/69	84x131	4	170x170
Wang 2016 [189]	IVIM	Breast care (Basel)	48	GE	3	8-channel	0, 10, 20, 50, 100, 200, 300, 400, 600, 800	2000/85	128x128	6	280x280
Bougias 2016 [185]	ADC	British Journal of Radiology	49	Siemens	1.5	4-channel	0, 1000	4500-5000/71-73	160x224	4 or 5	(280-360) x (280-360)
Eghtedari 2016 [186]	ADC	Quantitative Imaging in Medicine and Surgery	86	GE	1.5	8-channel	0, 100, 750, 1500	10530/99	136x156	3	400x340
Cho 2016 [187]	ADC, IVIM	European Radiology	62	Siemens	3	7-element	0, 30, 70, 100, 150, 200, 300, 400, 500, 800	2000/103	108x128	4	
Dijkstra 2016 [188]	ADC, IVIM	Journal of Magnetic Resonance Imaging	120	Siemens	1.5		0, 50, 200, 500, 800, 1000	9300/91	192x384	4	170x340
Liu 2016 [189]	IVIM	European Radiology	36	Philips	1.5	4-channel	0, 10, 20, 30, 50, 70, 100, 150, 200, 400, 600, 100	5065/66	200x196	5	300x300
Sharma 2016 [190]	ADC	Frontiers in Oncology	259	Siemens	1.5	4-channel	0, 500, 1000	5000/87	128x128	4 or 5	(250-350) x (250-350)
Arponen 2016 [191]	ADC	European Radiology	33	Philips	3	7-element	0, 200, 400, 600, 800	5000/95		4	

Si 2016 [192]	ADC	Diagnostic Interventional Radiology	56	Siemens	3	4-channel	0, 800		8300/80	192x192	4	340x150
Jiang 2016 [59]	DTI	Medical Science Monitor	56	Siemens	1.5		0, 1000		6900/90	144x192	5	380x285
Pawlik 2016 [193]	ADC	Polish Journal of Pathology	160	Siemens	1.5	4-channel	0, 100, 750, 1000		8300/83	192x100		200x400
Akin 2016 [195]	ADC	Journal of British Health	15	Siemens	3	16-channel	50, 500, 800		8300/93	192x192	4	(300-320) x (300-320)
Aribal 2016 [196]	ADC	European Journal of Radiology	129	Siemens	3	16-channel	50, 400, 800		9700/86		3	300x300
Osman 2016 [197]	ADC	Egyptian Journal of Radiology and Nuclear Medicine	56	Philips	1.5	Surface coil	0, 400, 800		10036/80	256x256	3	421x421
Fusco 2015 [198]	IVIM	BioMed Research International	26	Siemens	1.5		0, 50, 100, 150, 400, 800, 1000		7700/129	120x236	4	183x360
Sun 2015 [199]	ADC	Radiology	97	Siemens	1.5	4-channel	50, 1000		5900/80	192x192	4	384x384
Min 2015 [200]	ADC	World Journal of Surgical Oncology	52	GE	1.5		0, 400, 600, 800		9355/78.5	128x128	5	280x280
Kul 2015 [201]	ADC	Diagnostic Interventional Radiology	141	Siemens	1.5		50, 400, 1000		5400/94	192x192	3	326x326
Caivano 2015 [202]	ADC	Cancer Investigation	110	Philips	3	Surface coil	0, 750		10648/44	172x170	3	340x340
Cabuk 2015 [203]	ADC	Journal of Medical Imaging and Radiation Oncology	60	GE	1.5	4-channel	0, 200, 600, 1000		4450/87.5	128x128	4	340x420
Iima 2015 [176]	ADC, IVIM	Investigative Radiology	22	Siemens	3	16-channel	3, 5, 10, 20, 30, 50, 70, 100, 200, 400, 600, 800, 1000, 1500, 2000, 2500		4600/86	80x166	3	160x300
Ibrahim 2015 [204]	ADC	Egyptian Journal of Radiology and Nuclear Medicine	40	Philips	1.5		0, 400, 800		1036/80	256x256	3	350x350
Abowarda 2015 [205]	ADC	Egyptian Journal of Radiology and Nuclear Medicine	48	GE	1.5	4-channel	0, 750		7000/71.5	192x192	5	360x360
Wahab 2015 [206]	ADC	Egyptian Journal of Radiology and Nuclear Medicine	50	Toshiba	1.5	8-channel	0, 750		8200/95		3.5	
El Fiki 2015 [207]	ADC	Egyptian Journal of Radiology and Nuclear Medicine	30	Philips	1.5		0, 1000		7000/71.5	240x240	3	340x340
Altay 2014 [208]	ADC	JBR-BTR	37	Philips	1.5	4-channel	0, 1000		4200/95	256x256	5	230x230
Moukhtar 2014 [209]	ADC	Egyptian Journal of Radiology and Nuclear Medicine	80	GE	1.5	4-channel	0, 750		8400/93.8	128x128	4	340x340
Kul 2014 [210]	ADC	Journal of Magnetic Resonance Imaging	267	Siemens	1.5	2-channel	50, 400, 1000		5400/94	192x192	3	

Yoo 2014 [211]	ADC	Magnetic Resonance Imaging	89	Siemens	1.5	16-channel	0, 100, 500, 800, 1000	8500/80	108x220	3	340x340
Cai 2014 [212]	ADC	BMC Cancer	93	Siemens	3	Surface coil	0, 800	5400/86	192x192	5	300x300
Spick 2014 [213]	ADC	European Radiology	101	Siemens	1.5		50, 400, 800	6300/104		4	
Nogueira 2014 [215]	ADC	Clinical Radiology	53	Siemens	3	4-channel	50, 200, 400, 600, 800, 1000, 2000, 3000	4900/108	84x128	5	250x250
Bokacheva 2014 [216]	ADC, IVIM	Journal of Magnetic Resonance Imaging	35	GE	3	16-channel	0, 30, 60, 90, 120, 400, (450 in 7 cases), 600, 800, 1000	6000/56.4-120.7	98x98 (n=15), 128x128 (n=20)	5	(280-380) x (280-380)
Tan 2014 [217]	ADC	Clinical Radiology	50	GE	3		0, 500, 1000	8000/70.9-80.9	80x128		400x400
Hassan 2013 [218]	ADC	Alexandria Journal of Medicine	71	Siemens	1.5	4-channel	0, 400, 800	4800/98	192x192	4	(270-340) x (270-340)
Liu 2013 [219]	ADC, IVIM	European Journal of Radiology	84	Philips	1.5	16-channel	0, 10, 20, 30, 50, 70, 100, 150, 200, 400, 600, 1000	5065/66	200x196	5	300x300
Cakir 2013 [60]	DTI	European Journal of Radiology	52	Philips	3	4-channel	0, 1000	10461/55	165x96	2	(320-390) x (320-390)
Cheng 2013 [220]	ADC	Tumour Biology	174	GE	1.5	4-channel	0, 1000	8400/93.8	128x128	4	340x340
Oehi 2013 [221]	ADC	Breast Cancer	219	GE	1.5	8-channel	0, 1500	6900/75.5	128x128	5	360x360
Eyal 2012 [183]	DTI	European Journal of Radiology	46	Siemens	3	4-channel or 7-channel	0, 700	10400/120		1.9-2.5	
Orguc 2012 [222]	ADC	Singapore Medical Journal	108	GE	1.5	8-channel	0, 600	7900/88.9	192x192	5	(360-400) x (360-400)
Chen 2012 [223]	ADC	BMC Cancer	52	GE	1.5	8-channel	0, 600, 800, 1000	5000/72.1	128x128	5	320x320
Zhang 2012 [225]	ADC	Acta Radiologica	60	Philips	3	8-channel	0, 800	9000/67	140x256	2	266x373
Sonmez 2011 [226]	ADC	Wiener Klinische Wochenschrift	41	Siemens	1.5		0, 50, 200, 500, 1000	5800/90	256x256	5	
Inoue 2011 [227]	ADC	Japanese Journal of Radiology	105	Siemens	1.5		0, 250, 500, 750, 1000	5700/89	128x128	8	320x320
Baltzer 2011 [61]	DTI	European Radiology	59	Siemens	1.5	4-channel	0, 1000	8800/139		5	
Kul 2011 [228]	ADC	American Journal of Roentgenology	84	Siemens	1.5		50, 400, 1000	4000/94	192x192	3	
Fornasa 2011 [229]	ADC	La Radiologica Medica	78	GE	1.5	Surface coil	0, 800	4000/55.2	64x64	7	340x340

Gouhar 2011 [230]	ADC	Egyptian Journal of Radiology and Nuclear Medicine	62	GE	1.5	16-channel	0, 1000	3750/72.6	3	330x330
Imamura 2010 [231]	ADC	Magnetic Resonance in Medical Sciences	27	Philips	1.5	Surface coil	0, 1000	4508/67	8	420x420
Jin 2010 [232]	ADC	Academic Radiology	91	GE	1.5	4-channel	0, 600, 1000	4100/60	4	320-360
Partridge 2010 [233]	ADC	Journal of Magnetic Resonance Imaging	86	GE	1.5	8-channel	0, 600	7000/71.5	5	360x360
Belli 2010 [234]	ADC	La Radiologica Medica	21	GE	1.5		0, 1000	5150/minimum	4	(320-340) x (320-340)
Yabuuchi 2010 [235]	ADC	European Journal of Radiology	70	Philips	1.5	4-channel	0, 500, 1000	6839/51	5	300x300
Palle 2009 [236]	ADC	Indian Journal of Radiological Imaging	200	Siemens	1.5		0, 500, 1000	1800/75	3	350x350
Pereira 2009 [237]	ADC	American Journal of Roentgenology	45	GE	1.5	8-channel	0, 250, 500, 750, 1000	1800/93.8	5	360x360
Tozaki 2009 [238]	ADC	American Journal of Roentgenology	165	Siemens	1.5	4-channel	0, 500, 1000, 1500, 2000, 3000	8000/96	3	330x330
Lo 2009 [239]	ADC	Journal of Computer Assisted Tomography	31	Siemens	3	4-channel	0, 1000	3900/78	6	350x350
Yabuuchi 2008 [240]	ADC	Journal of Magnetic Resonance Imaging	71	Philips	1.5	4-channel	0, 500, 1000	6839/51	5	300x300
Hatakenaka 2008 [40]	ADC	Magnetic Resonance in Medical Sciences	136	Siemens	1.5		0, 500, 1000	2000-4000/100-135	5	(240-300) x (240-300)
Marini 2007 [44]	ADC	European Radiology	60	Siemens	1.5		0, 1000	2690/79	4	300x177
Wenkel 2007 [241]	ADC	Academic Radiology	51	Siemens	1.5		50, 400, 800	8400/98	4	340x170
Rubesova 2006 [242]	ADC	Journal of Magnetic Resonance Imaging	78	Siemens	1.5		0, 200, 400, 600, 1000	2600/110	4	160x160
Woodhams 2005 [243]	ADC	Magnetic Resonance in Medical Sciences	190	GE	1.5	Surface coil	0, 750	5000/61.8	6	320x240
Guo 2002 [39]	ADC	Journal of Magnetic Resonance Imaging	52	GE	1.5	Surface coil	0, 250, 500, 750, 1000 (n=10), 0, 100 (n=42)	unknown / 99	5	240x240

Appendix Table 1 Details of studies included in the meta-analysis

n = number of patients, TR = repetition time, TE = echo time, FOV = field-of-view, ADC = apparent diffusion coefficient, DTI = diffusion tensor imaging, IVIM = intra-voxel incoherent motion, EPI = echo-planar imaging

Appendix 2

S+H	S+H+T	S+H+P	S+H+T+P
Flatness	Elongation	Group	Group
LeastAxisLength	Flatness	Grade	
Maximum2DDiameterColumn	LeastAxisLength	Age	
Sphericity	Maximum2DDiameterColumn	LeastAxisLength	
SurfaceVolumeRatio	Maximum2DDiameterRow	Maximum2DDiameterColumn	
10Percentile_1st	Sphericity	Sphericity	
Kurtosis_1st	InterquartileRange_1st	SurfaceArea	
Skewness_1st	Kurtosis_1st	Kurtosis_1st	
InterquartileRange_2nd	ClusterProminence_1st	InterquartileRange_2nd	
Kurtosis_2nd	ClusterShade_1st	Skewness_2nd	
Skewness_2nd	Imc2_1st	Kurtosis_3rd	
10Percentile_3rd	MCC_1st	Minimum_3rd	
Kurtosis_3rd	LargeAreaHighGrayLevelEmphasis_1st	Uniformity_3rd	
Minimum_3rd	ZoneEntropy_1st	InterquartileRange_4th	
Uniformity_3rd	Kurtosis_2nd	Variance_4th	
InterquartileRange_4th	Skewness_2nd	Kurtosis_5th	
Skewness_4th	ClusterProminence_2nd	Median_5th	
Kurtosis_5th	Imc2_2nd	Skewness_5th	
Uniformity_5th	LowGrayLevelZoneEmphasis_2nd	Uniformity_5th	
	SmallAreaHighGrayLevelEmphasis_2nd		
	SmallAreaLowGrayLevelEmphasis_2nd		
	Kurtosis_3rd		
	Maximum_3rd		
	Minimum_3rd		
	Idmn_3rd		

	LongRunLowGrayLevelEmphas_3rd		
	LargeAreaHighGrayLevelEmphas_3rd		
	InterquartileRange_4th		
	Kurtosis_4th		
	Skewness_4th		
	ClusterShade_4th		
	InverseVariance_4th		
	MCC_4th		
	GrayLevelNonUniformity_4th		
	SizeZoneNonUniformity_4th		
	SmallAreaEmphasis_4th		
	SmallAreaLowGrayLevelEmphas_4th		
	ZoneEntropy_4th		
	Kurtosis_5th		
	Correlation_5th		
	Idmn_5th		
	Imc1_5th		
	MCC_5th		
	MaximumProbability_5th		
	HighGrayLevelZoneEmphasis_5th		
	SizeZoneNonUniformityNormalized_5th		
	SmallAreaLowGrayLevelEmphas_5th		

Reduced feature sets

S+H	S+H+T	S+H+P	S+H+T+P
LeastAxisLength	Flatness	Group	Group
Maximum2DDiameterColumn	LeastAxisLength	Grade	Grade
Sphericity	Maximum2DDiameterColumn	Age	Age

SurfaceVolumeRatio	Sphericity	LeastAxisLength	LeastAxisLength
10Percentile_1st	Kurtosis_1st	Maximum2DDiameterColumn	Maximum2DDiameterColumn
Kurtosis_1st	ClusterShade_1st	Sphericity	Sphericity
Skewness_1st	Imc2_1st	SurfaceVolumeRatio	Kurtosis_1st
Skewness_2nd	MCC_1st	10Percentile_1st	Minimum_1st
10Percentile_3rd	LargeAreaLowGrayLevelEmphasis_1st	Kurtosis_1st	Correlation_1st
Kurtosis_3rd	ZoneEntropy_1st	Skewness_2nd	Imc2_1st
Variance_3rd	Kurtosis_2nd	Kurtosis_3rd	RunVariance_1st
MeanAbsoluteDeviation_4th	Skewness_2nd	Variance_3rd	SmallAreaHighGrayLevelEmphasis_1st
Skewness_4th	Correlation_2nd	MeanAbsoluteDeviation_4th	Skewness_2nd
Kurtosis_5th	LongRunLowGrayLevelEmphasis_2nd	Skewness_4th	Correlation_2nd
	10Percentile_3rd	Kurtosis_5th	LongRunLowGrayLevelEmphasis_2nd
	Kurtosis_3rd	Minimum_5th	LargeAreaHighGrayLevelEmphasis_2nd
	Correlation_3rd	Skewness_5th	Kurtosis_3rd
	LongRunLowGrayLevelEmphasis_3rd		ClusterShade_3rd
	LargeAreaHighGrayLevelEmphasis_3rd		Correlation_3rd
	Kurtosis_4th		LongRunLowGrayLevelEmphasis_3rd
	Skewness_4th		LargeAreaHighGrayLevelEmphasis_3rd
	ClusterShade_4th		Imc1_4th
	Imc1_4th		MCC_4th
	MCC_4th		SmallAreaEmphasis_4th
	SizeZoneNonUniformity_4th		Kurtosis_5th
	SmallAreaEmphasis_4th		Skewness_5th
	Kurtosis_5th		ClusterShade_5th
	MCC_5th		MCC_5th
	SizeZoneNonUniformityNormalized_5th		SizeZoneNonUniformityNormalized_5th
	SmallAreaHighGrayLevelEmphasis_5th		SmallAreaHighGrayLevelEmphasis_5th

Time points: Image-derived features

1st post-contrast	2nd post-contrast	3rd post-contrast	4th post-contrast	5th post-contrast
Energy_1st	10Percentile_2nd	10Percentile_3rd	InterquartileRange_4th	10Percentile_5th
InterquartileRange_1st	InterquartileRange_2nd	Kurtosis_3rd	Kurtosis_4th	Kurtosis_5th
Maximum_1st	Kurtosis_2nd	RobustMeanAbsoluteDeviation_3rd	Mean_4th	Skewness_5th
Median_1st	Skewness_2nd	Skewness_3rd	Minimum_4th	Flatness_5th
Range_1st	TotalEnergy_2nd	Flatness_3rd	Skewness_4th	LeastAxisLength_5th
Skewness_1st	Flatness_2nd	LeastAxisLength_3rd	Flatness_4th	Maximum2DDiameterColumn_5th
TotalEnergy_1st	LeastAxisLength_2nd	Maximum2DDiameterColumn_3rd	LeastAxisLength_4th	MinorAxisLength_5th
Flatness_1st	Maximum2DDiameterColumn_2nd	MinorAxisLength_3rd	Maximum2DDiameterColumn_4th	Sphericity_5th
LeastAxisLength_1st	Sphericity_2nd	Sphericity_3rd	Maximum2DDiameterRow_4th	SurfaceArea_5th
MajorAxisLength_1st	SurfaceArea_2nd	ClusterShade_3rd	Sphericity_4th	Correlation_5th
Maximum2DDiameterColumn_1st	Correlation_2nd	ClusterTendency_3rd	SurfaceArea_4th	Imc1_5th
Sphericity_1st	DifferenceEntropy_2nd	Correlation_3rd	ClusterShade_4th	MCC_5th
SurfaceArea_1st	Imc1_2nd	Imc1_3rd	ClusterTendency_4th	MaximumProbability_5th
Correlation_1st	Imc2_2nd	MCC_3rd	Contrast_4th	LargeAreaLowGrayLevelEmphasis_5th
DifferenceVariance_1st	MCC_2nd	LongRunLowGrayLevelEmphasis_3rd	Correlation_4th	LowGrayLevelZoneEmphasis_5th
Idmn_1st	MaximumProbability_2nd	LargeAreaHighGrayLevelEmphasis_3rd	Idmn_4th	SizeZoneNonUniformity_5th
Imc2_1st	LargeAreaHighGrayLevelEmphasis_2nd	LargeAreaLowGrayLevelEmphasis_3rd	Imc1_4th	SmallAreaEmphasis_5th
JointEntropy_1st	LowGrayLevelZoneEmphasis_2nd	LowGrayLevelZoneEmphasis_3rd	MCC_4th	SmallAreaHighGrayLevelEmphasis_5th
MCC_1st	SizeZoneNonUniformity_2nd	SizeZoneNonUniformity_3rd	GrayLevelNonUniformity_4th	
MaximumProbability_1st			LongRunLowGrayLevelEmphasis_4th	

GrayLevelNonUniformity_1st			LargeAreaHighGrayLevelEmphasis_4th	
ShortRunLowGrayLevelEmphasis_1st			LargeAreaLowGrayLevelEmphasis_4th	
GrayLevelNonUniformityNormalized_1st			SizeZoneNonUniformity_4th	
LargeAreaHighGrayLevelEmphasis_1st			SmallAreaEmphasis_4th	
LargeAreaLowGrayLevelEmphasis_1st			SmallAreaLowGrayLevelEmphasis_4th	
ZoneEntropy_1st			ZoneEntropy_4th	

Time points: Image-derived + clinicopathological features

1st post-contrast	2nd post-contrast	3rd post-contrast	4th post-contrast	5th post-contrast
Group	Group	Group	Group	Group
	Grade	Age		
	Skewness_2nd	Subtype		
	ClusterProminence_2nd	10Percentile_3rd		
	LargeAreaHighGrayLevelEmphasis_2nd	Energy_3rd		
		Kurtosis_3rd		
		Maximum_3rd		
		Minimum_3rd		
		RobustMeanAbsoluteDeviation_3rd		
		Skewness_3rd		
		TotalEnergy_3rd		
		Flatness_3rd		
		LeastAxisLength_3rd		
		MajorAxisLength_3rd		
		Maximum2DDiameterColumn_3rd		
		Maximum2DDiameterSlice_3rd		
		Sphericity_3rd		
		SurfaceArea_3rd		
		ClusterShade_3rd		

		Correlation_3rd		
		Imc1_3rd		
		Imc2_3rd		
		MCC_3rd		
		LargeAreaHighGrayLevelEmphasis_3rd		
		LowGrayLevelZoneEmphasis_3rd		
		SizeZoneNonUniformity_3rd		

Time points: Added value of time points

1st post-contrast	1+2	1+2+3	1+2+3+4	S+H+T+P
Group	Group	Group	Group	Group
	Grade	Grade		Grade
	Skewness_2nd	LargeAreaHighGrayLevelEmphasis_2nd		Age
	ClusterProminence_2nd	Minimum_3rd		LeastAxisLength
	LargeAreaHighGrayLevelEmphasis_2nd	ClusterProminence_3rd		Maximum2DDiameterColumn
		ClusterTendency_3rd		Sphericity
				SurfaceVolumeRatio
				10Percentile_1st
				Kurtosis_1st
				Skewness_2nd
				Kurtosis_3rd
				Variance_3rd
				MeanAbsoluteDeviation_4th
				Skewness_4th
				Kurtosis_5th
				Minimum_5th
				Skewness_5th

2378

Quantitative Bilateral Sodium and Proton Breast Imaging at 3T

Gabrielle C Baxter¹, Mary A McLean², Joshua D Kaggie¹, Frank Riemer³, Ferdia A Gallagher¹, Andrew D James^{4,5}, Aneurin J Kennerley^{5,6}, Rolf F Schulte⁷, William J Brackenbury^{4,5}, and Fiona J Gilbert¹

¹Department of Radiology, University of Cambridge, Cambridge, United Kingdom, ²Cancer Research UK, Cambridge, United Kingdom, ³Department of Radiology, Haukeland University Hospital, Bergen, Norway,

⁴Department of Biology, University of York, York, United Kingdom, ⁵York Biomedical Research Institute, University of York, York, United Kingdom, ⁶Department of Chemistry, University of York, York, United Kingdom, ⁷GE Healthcare, Munich, Germany

Synopsis

Sodium (²³Na)-MRI was performed using a dual-tuned bilateral ²³Na/¹H breast coil on four healthy volunteers. Images acquired using ²³Na-MRI show a relationship between sodium concentration and diffusion properties of breast tissue. Sodium signal variations due to mild flip angle non-uniformity were observed between left and right breasts. After correction using B₁ mapping, tissue sodium concentration (TSC) maps were more closely matched between breasts.

Introduction

Sodium (²³Na)-MRI is an emerging technique that is posited as enabling the differentiation of malignant and benign breast lesions and normal fibroglandular tissue based on increases in total tissue sodium concentration (TSC)¹. ²³Na-MRI, alongside other multiparametric proton (¹H)-MRI techniques such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced imaging (DCE-MRI), can provide complementary information about the physiological and biochemical state of tumours. Furthermore, as changes in sodium concentration are likely to occur before changes in cellularity or vascularity as measured by DWI and DCE-MRI, ²³Na-MRI may provide more immediate information about changes in tumour physiology to assess the effects of therapy².

Methods

Four healthy volunteers (mean age 35.5 years, range 25 – 52 years) were scanned with informed consent and the approval of the local ethics board on a 3T system (MR750, GE Healthcare, Waukesha, WI) using a dual-tuned bilateral ²³Na/¹H breast coil (Rapid Biomedical, Rimpf, Germany) in the prone position. The coil consisted of 4 transmit/receive sodium channels and 16 receive-only proton channels.

²³Na-MRI protocol: ²³Na-MRI was performed using a 3D cones ultra-short echo time trajectory³ with TR/TE = 100/0.46ms, flip angle = 90 degrees, voxel size = 3x3x6mm³ for a 36 x 36 cm² field-of-view (FOV), number of averages = 4, interleaves = 1402, total scan time = 9 minutes and 21 seconds. Inversion recovery images were also acquired with inversion time 30ms, TR 250ms, 2 averages. Low resolution sodium images were acquired at flip angles of 40 and 80 degrees for B₁ mapping. For calibration purposes, a fiducial with sodium concentration 78.5mM was placed in the FOV.

¹H-MRI protocol: A 3D T₁-weighted fast spoiled gradient echo sequence (TR/TE = 4.42/1.97ms, voxel size = 1.1 x 1.6 x 2.8mm³, FOV = 36 x 36 cm², scan time = 1 minute 27 seconds) was performed to generate water and fat images. DWI was performed using a single-shot echo-planar imaging sequence with TR/TE = 4000/94.2ms, voxel size = 2.8 x 2.8 x 4mm³, FOV = 36 x 36 cm², acceleration factor = 2, averages = 4, b-values = 0, 100, 500, 1000, 1500, 2000, 2500 s/mm², total scan time = 10 minutes 12 seconds.

Image Processing: ²³Na-MRI images were obtained via a re-gridding routine before Fourier transformation. B₁ maps were generated using the double angle method⁴ as the ratio of the actual to nominal flip angle. Regions of interest were drawn on B₁ maps covering two thirds of left and right breasts to measure flip angle variation between breasts. TSC maps were generated based on the calibration fiducial concentration, assuming that the TR was sufficiently long to avoid T₁ saturation. Given the distance of the fiducial from the breast, TSC maps were corrected with B₁ maps using the spoiled echo gradient signal equation. ADC maps were generated using in-house software developed in MATLAB (version 2018b).

Results

Figure 1 shows ²³Na-MRI and T₁-weighted water and fat images from two volunteers. Qualitatively, there was more sodium signal in fibroglandular tissue compared to fatty tissue. For both volunteers, sodium signal was higher in the right breast than the left, which is likely a systematic error due to a 10-20% higher B₁ field in the left versus right coil.

An example of a B₁ map and a B₁ corrected TSC map are shown in Figure 2. There was significant variation in the B₁ map in the left/right directions and anterior/posterior directions. There was a higher flip angle observed in the left breast. The mean flip angle variation between left and right breasts measured across all patients was 24 ± 11 degrees. After flip angle correction, the TSC matched more closely between breast sides (Figure 2C), although this correction does not account for receive sensitivities. High TSC was observed in the heart, which is explainable due to large sodium concentrations in the blood pool.

Figure 3 shows a comparison of ²³Na-MRI and DWI. Similar anatomical structures were observed in both types of image. Areas with high sodium signal correspond to areas with high ADC.

Discussion

High quality sodium and proton images were acquired using a dual-tuned bilateral coil with a unique number of sodium and proton channels. The mean variation in flip angle between left and right breasts was comparable to previously published values for ¹H-MRI breast imaging⁵. The sodium concentrations measured in fibroglandular tissue were slightly higher than values published previously¹. However, given that our study only included healthy volunteers, it is not clear whether these values are a suitable reference standard.

Similar anatomical structures were observed in ²³Na-MRI and DWI, suggesting a relationship between tissue sodium concentration and diffusion properties. Future work will include DCE-MRI to perform quantitative pharmacokinetic modelling, aiming to investigate the relationship of tumour vascularity to sodium concentration in breast tissue.

Conclusion

Images acquired using ²³Na-MRI show a relationship between sodium concentration, anatomical structure and diffusion properties. B₁ correction is required for accurate quantification of TSC.

Acknowledgements

This work was supported by GlaxoSmithKline, Cancer Research UK and the National Institute of Health Research Cambridge Biomedical Research Centre.

References

1. Ouwerkerk, R. et al. Elevated tissue sodium concentration in malignant breast lesions detected with non-invasive ^{23}Na MRI. *Breast Cancer Res. Treat.* 106, 151–160 (2007).
2. Madelin, G., Lee, J.-S., Regatte, R. R. & Jerschow, A. Sodium MRI: methods and applications. *Prog. Nucl. Magn. Reson. Spectrosc.* 79, 14–47 (2014).
3. Gurney, P. T., Hargreaves, B. A. & Nishimura, D. G. Design and analysis of a practical 3D cones trajectory. *Magn. Reson. Med.* 55, 575–582 (2006).
4. Cunningham, C. H., Pauly, J. M. & Nayak, K. S. Saturated double-angle method for rapid B_1+ mapping. *Magn. Reson. Med.* 55, 1326–33 (2006).
5. Azlan, C. A. et al. B_1 transmission-field inhomogeneity and enhancement ratio errors in dynamic contrast-enhanced MRI (DCE-MRI) of the breast at 3T. *J. Magn. Reson. Imaging* 31, 234–9 (2010).

Figures

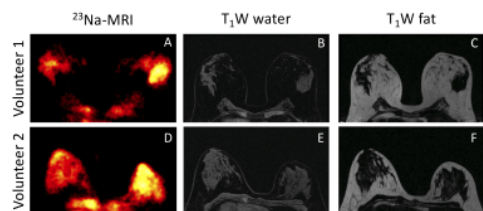


Figure 1. ^{23}Na -MRI (A, D) and T_1 -weighted water (B,E) and fat images (C,F).

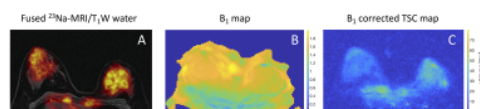


Figure 2. A) Fused ^{23}Na -MRI and T_1 -weighted water images, B) B_1 map and C) B_1 corrected tissue sodium concentration (TSC) map.



Figure 3. A) ^{23}Na -MRI, B) DWI with $b = 1000 \text{ s/mm}^2$ and C) Apparent Diffusion Coefficient (ADC) map.