# DEVELOPMENT OF NOVEL METHODS FOR OBTAINING ROBUST DYNAMIC SUSCEPTIBILITY CONTRAST MAGNETIC RESONANCE IMAGING BIOMARKERS FROM DISEASED BRAIN IN CHILDREN

By

Stephen Jake Powell

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Physical Sciences for Health
School of Chemistry
College of Engineering and Sciences
University of Birmingham

March 2021

# Abstract

Dynamic susceptibility contrast (DSC-) MRI is an important imaging technique from which estimates of perfusion measures including cerebral blood volume (CBV), cerebral blood flow (CBF) and mean transit time (MTT) can be calculated. These perfusion measures can be used to indicate health in a range of diseases. However, acquisition protocol varies from centre-to-centre, which leads to variability in data quality between centres and limits the clinical applicability of DSC-MRI. Currently, the recommended process for assessing data quality is by eye, which is very time consuming and subjective between reviewers.

In this work an automated processing pipeline for DSC-MRI was produced. Work to develop the pipeline demonstrated that data quality of DSC-MRI data can be assessed using machine learning classifiers, which were trained using metrics calculated from the data and the results of qualitative review. It also showed that it was possible to denoise the data using singular value decomposition (SVD) based methods, which were validated on a simulator and confirmed in patient data. The pipeline created was applied to a multicentre patient dataset where it demonstrated the importance of denoising DSC-MRI data in improving data quality and how data quality can vary with acquisition protocol. It was also applied to a single centre study of patients receiving differing treatments for brain tumours and suggested there are no significant changes in relative CBV (rCBV) in non-tumour brain between differing treatment groups. The pipeline developed during this work has wider applications in other imaging modalities and could be adapted to be applied to other perfusion imaging methods, such as dynamic contrast enhanced (DCE-) MRI, or any other imaging modality that involves analysis of a signal variation with time, such as computed tomography (CT) perfusion imaging or positron emission tomography (PET).

# Declaration

I declare that the work presented in this thesis is entirely my own with the following exceptions:

- Clinical data from the CCLG database was acquired by radiographers from centres across the UK. In this thesis data from four different centres was used.
- In Chapter 3, qualitative review of the subset of signal-time courses was carried out by Dr Stephanie Withey.
- In Chapter 6, grey matter and white matter regions of interest were also evaluated by Professor Andrew Peet and Dr Stephanie Withey.
- K2 values produced from leakage correction and presented in the appendix were calculated using an in-house tool developed by Dr Stephanie Withey.

## Acknowledgements

I would like to thank my supervisors Professor Andrew Peet, Dr Stephanie Withey and Dr Yu Sun, and also Dr James Grist for their support and guidance throughout my PhD. Thank you for the opportunity to carry out research in a really interesting area of MRI.

I would also like to thank everyone from the Children's Brain Tumour Team, I have really enjoyed my time working in the group. Thanks to Emma Metcalfe-Smith, Dadi Zhao and Sonal Nandanwar for sharing an office with me and for all the tea breaks. Thanks to Physical Sciences for Health Centre for Doctoral Training, it has been a great community of scientists to be a part of.

Finally, I would like to thank my family for all their support through 8 years of university. Thanks to my friends for the weekly skype chats, which made writing up in a lockdown a lot more bearable!

This work was funded by the engineering and physical sciences research council (EPSRC) through a studentship from the Sci-Phy-4-Health CDT (EP/L016346/1). Support was also received from the National Institute for Health Research (NIHR) via a research professorship (RP-R2-12-019), the Children's Research Fund, Help Harry Help Others (HHHO), Cancer Research UK, the Experimental Cancer Medicine Centre Pediatric Network (C8232/A25261), the Children's Cancer Fund, and the Little Princess Trust.

## Conference Presentations

- Poster presentation, ISMRM, Virtual 2020, An Automated Processing Pipeline to Assess and Improve Data Quality for Multicentre Pediatric Dynamic Susceptibility Contrast Imaging
- Poster presentation, BICISMRM, Sheffield 2019, Comparing Denoising Methods for Dynamic Susceptibility Contrast (DSC-) MRI Using Simulations
- Poster presentation, ISMRM, Montreal 2019, An Algorithm for the Automated Quality Assessment and Perfusion Biomarker Determination of Multicentre Dynamic Susceptibility Contrast (DSC-) DSC MRI
- Power pitch presentation, PG BICISMRM, Birmingham 2019, An Automated Pipeline for Quality Assessment and Biomarker Determination of Multicentre Dynamic Susceptibility Contrast (DSC-) MRI
- Poster presentation, BICISMRM, Oxford 2018, A Dynamic Susceptibility Contrast MRI Simulator for Determining Biomarker Accuracy
- Power pitch presentation, PG BICISMRM, London 2018, Development of an Automated Quality Control Method for Dynamic Susceptibility Contrast (DSC-) MRI

# Table of Contents

# List of Illustrations

# List of Tables

# List of Abbreviations

**A**

AIF = Arterial Input Function

ANOVA = Analysis of Variance

ASFNR = American Society of Functional Neuroradiology

ASL = Arterial Spin Labelling

AT = Arrival Time

AUC = Area Under Curve

**B**

BBB = Blood Brain Barrier

**C**

CBF = Cerebral Blood Flow

CBV = Cerebral Blood Volume

CMRO = Cerebral Metabolic Rate of Oxygen

CNS = Central Nervous System

CS = Craniospinal

CSF = Cerebrospinal Fluid

CSI = Chemical Shift Imaging

CT = Computed Tomography

CTC = Concentration Time Curve

CV = Cross Validation

**D**

DCE = Dynamic Contrast Enhanced

DSC = Dynamic Susceptibility Contrast

DWI = Diffusion Weighted Imaging

**E**

EES = Extravascular Extracellular Space

EIBALL = European Imaging Biomarkers Alliance

EPI = Echo Planar Imaging

**F**

FA = Flip Angle

FFRT = Focal Fractionated Radiotherapy

FID = Free Induction Decay

FOV = Field of View

FWHM = Full Width Half Maximum

**G**

GBM = Glioblastoma Multiforme

Gd-DTPA = Gadolinium diethylene-triaminepenta-acetic acid

GE = Gradient Echo

GE-EPI = Gradient Echo – Echo Planar Imaging

GM = Grey Matter

GRAPPA = Generalised Autocalibrating Partially Parallel Acquisition

**H**

HGG = High Grade Glioma

HOSVD = Higher Order Singular Value Decomposition

**I**

IDH = Isocitrate Dehydrogenase

IVIM = Intravoxel Incoherent Motion

**M**

MRI = Magnetic Resonance Imaging

MRS = Magnetic Resonance Spectroscopy

MSD = Maximum Signal Drop

MTT = Mean Transit Time

**N**

nCBV = Normalised Cerebral Blood Volume

NEMA = National Electrical Manufacturers Association

NIRS = Near Infrared Spectroscopy

NMR = Nuclear Magnetic Resonance

**P**

PC = Principal Component

PCA = Principal Component Analysis

PD = Proton Density

PET = Positron Emission Tomography

PSR = Percentage Signal Recovery

PVE = Partial Volume Effect

**Q**

QIB = Quantitative Imaging Biomarker

QIBA = Quantitative Imaging Biomarkers Alliance

**R**

RADS = Reporting and Data Systems

rCBV = Relative Cerebral Blood Volume

RF = Radiofrequency

RMSE = Root Mean Square Error

ROC = Receiver Operator Curve

ROI = Region of Interest

RSNA = Radiological Society of North America

**S**

SDNR = Signal Drop to Noise Ratio

SENSE = Sensitivity Encoded

SNR = Signal-to-noise Ratio

sPRESTO = SENSE Principles of Echo-Shifting with a Train of Observations

SPSP = Spatial-Spectral

SRS = Stereotactic Radiosurgery

SURE = Stein’s Unbiased Risk Estimator

SVD = Singular Value Decomposition

SVM = Support Vector Machine

**T**

TE = Echo Time

TIC = Time Intensity Curve

TR = Repetition Time

TTP = Time to Peak

**W**

WM = White Matter

# Chapter 1: Introduction

# 1. Introduction

## 1.1. Perfusion

Perfusion is a physiological process, which describes the delivery of blood to tissues and organs via the capillaries. It is measured as volume of blood per unit of tissue (1). The process of perfusion is used to deliver oxygen and nutrients to the tissue and is also responsible for the removal of waste products (2). Therefore, perfusion is crucial to tissue health and changes in perfusion can be associated with a variety of pathologies. Measurement of perfusion can lead to further understanding of normal and pathologic physiologies and is an important tool in disease diagnosis and treatment (3). Consequently, ensuring perfusion measurements are quantified is an important step in applying them as quantitative imaging biomarkers (4).

Examples of applications of perfusion measurement in disease diagnosis include coronary artery disease, ischaemia, tumours, abscesses, and neurodegenerative diseases (5, 6). Examples of application of perfusion measurement in disease treatment include: assessing the performance of drugs used to treat coronary artery disease and brain tumour patients (7, 8); monitoring treatment response in patients with Parkinson's disease (9); and assessing perfusion pre- and post-surgical treatment (10).

Perfusion can be characterised using a series of physiological parameters, as summarised in Figure 1.1. Tissue blood flow, $F_T$, is the blood flow per volume of tissue and is measured as mL $min^{-1}$ $100mL^{-1}$. In brain measurements this is often referred to as cerebral blood flow (CBF). The average time taken for the blood to pass through the network is defined as the mean transit time (MTT). Blood volume fraction, $\upsilon_b$, defines the volume of blood contained within a volume of tissue and is measured as mL $100mL^{-1}$. In brain measurements this is often referred to as cerebral blood volume (CBV). Extracellular and extravascular volume fraction, $\upsilon_e$, defines the extracellular extravascular volume within the tissue volume. The permeability-

surface area product, PS, defines the flow of blood through the capillary membrane into the extravascular extracellular space and is also measured as mL $min^{-1}$ $100mL^{-1}$. The leakiness of the capillaries can be assessed through the use of $K^{trans}$ which is a combination of PS and tissue blood flow (11).

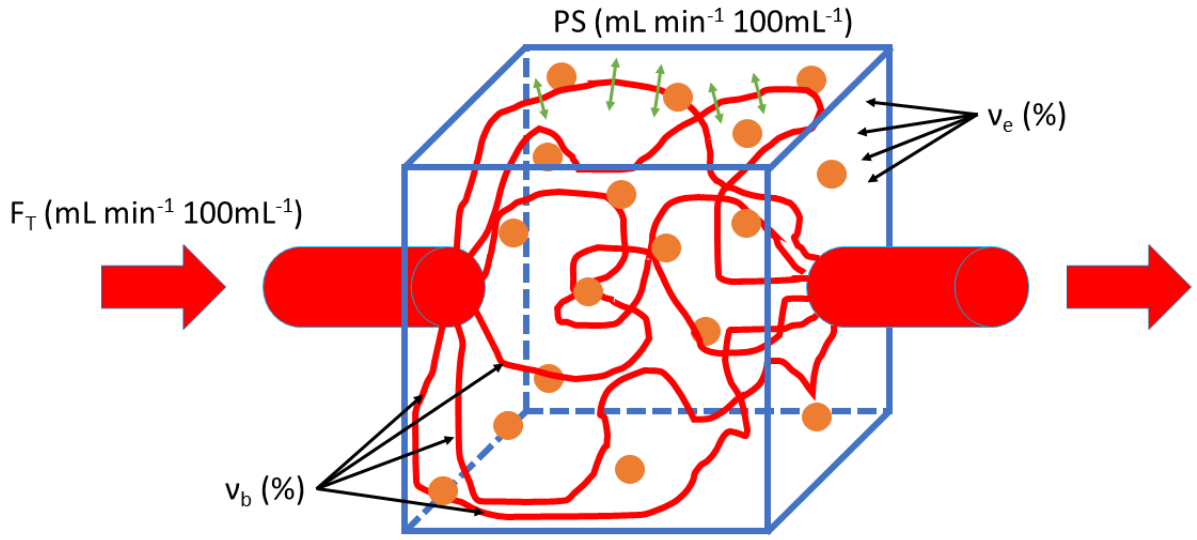



*Figure 1.1: Perfusion parameters that can be obtained from microcirculation. The blue cube represents a volume of tissue, the red tubes represent the capillaries, the orange circles represent cells, and the green arrows represent the leakage of blood from the capillaries into the extravascular extracellular space (characterised by PS).*

All of these parameters have applications in assessing different pathologies and estimates of their values can be made using perfusion imaging.

## 1.2. Perfusion Imaging

There are several different imaging techniques, which can be used to image perfusion within the body including: computed tomography (CT), positron emission tomography (PET), ultrasound, near infrared spectroscopy (NIRS), and magnetic resonance imaging (MRI).

In CT perfusion imaging an iodinated contrast agent is injected and CT images are acquired as the contrast agent passes through the body. From this method it is possible to obtain estimates of CBV, CBF and vascular MTT (12). CT perfusion imaging has quick image acquisition and

is a widely available imaging technique. However, it requires additional radiation doses compared to standard CT and there are risks associated with iodinated contrast agents (13).

In PET, a radioactive positron-emitting tracer is injected into the blood stream and the photons produced from the interactions between the emitted positrons and the surrounding tissue are detected using a PET camera (14). PET can be used to determine estimates of regional CBV, regional CBF and regional cerebral metabolic rate of oxygen (CMRO) (15). PET imaging has high sensitivity, can provide accurate estimation of tracer concentration, and has good diagnostic accuracy. However, it has limited SNR and spatial resolution, and requires access to a cyclotron to produce the tracer, which means it is not applicable in all settings (16).

Ultrasound uses the Doppler effect to provide estimates of blood flow. However, tissue motion can make measurements difficult, specifically for smaller vessels. Therefore, microbubbles can be used as a contrast agent which helps distinguish tissue from blood vessels (17). The use of a contrast agent means that a time intensity curve (TIC) can be produced from a region of interest (ROI). MTT and time to peak (TTP) can be calculated from the TIC along with several other parameters (18). Ultrasound has the advantage of being very repeatable and more portable than other methods as it can be performed at the bedside (17). However, it is susceptible to attenuation artefacts and has poor sensitivity to slow blood flow and capillary flow (17, 19). It is difficult to apply in the brain, as the skull blocks the ultrasound waves. However, it can be applied in neonates as the skull is not fully developed (20).

NIRS uses near infrared light to provide estimates of the oxygen saturation of the tissue, which can be used to estimate perfusion. Infrared light is transmitted into the tissue and is

detected by nearby receivers. Differences in intensity between the transmitted and received light are used to calculate estimates of oxygen saturation (21). NIRS provides a non-invasive measure of oxygen saturation. However, it has limited spatial resolution and limited penetration, meaning that it is only suitable for measurements within the cortex. The acquired signal can also be contaminated by contributions from other molecules (e.g. haemoglobin) (21, 22).

MRI can be used to provide estimates of tissue perfusion using specifically designed protocols, which have led to clinical applications in assessing perfusion related abnormalities (23). Perfusion MRI techniques can be split into two categories: those that use exogenous contrast agents and those that use endogenous contrast agents (24). Both dynamic susceptibility contrast (DSC-) MRI and dynamic contrast enhanced (DCE-) MRI use gadolinium based exogenous contrast agents, which are injected into the patient (25). Arterial spin labelling (ASL) uses magnetically labelled blood as an endogenous tracer (26) and intravoxel incoherent motion (IVIM) uses the pseudo-diffusion of blood water in the capillaries to estimate perfusion (27). Each of these MRI methods will be discussed in more detail in the sections below. Table 1.1 summarises all the imaging methods discussed in this chapter, the parameters obtained from them and their pros and cons. The cons column focuses on problems specific to each imaging technique. This is because problems such as noise, motion, and variability in acquisition protocol between centres, are common across multiple imaging techniques.

### 1.2.1. DSC-MRI

DSC-MRI is a technique used mostly within the brain, where the passage of a gadolinium based contrast agent is imaged using a dynamically acquired $T_2$ or $T_2^*$ weighted sequence (28). The contrast agent causes shortening in $T_1$, $T_2$, and $T_2^*$, therefore causing a decrease in

signal as it passes through the body (29). The equations for the change in $T_1$, $\Delta T_1$, and $T_2$, $\Delta T_2$, with contrast agent are shown in equations 1.1 and 1.2. In these equations $r_1$ and $r_2$ are the $T_1$ and $T_2$ relaxation rates respectively, and C is the concentration of contrast agent (30).

$$\frac{1}{\Delta T_1} = r_1 \cdot C \tag{1.1}$$

$$\frac{1}{\Delta T_2} = r_2 \cdot C \tag{1.2}$$

Therefore, for each pixel within a DSC-MRI image it is possible to plot a signal-time course which shows how the signal intensity changes with time. From the signal-time course, it is possible to calculate: the signal-to-noise ratio (SNR); the maximum signal drop (MSD); and the percentage signal recovery (PSR) (31, 32). The signal-time course can also be converted into a concentration time curve (CTC), from which estimates of perfusion parameters such as CBV, CBF and vascular MTT can be calculated. Further details on how these perfusion parameters are calculated can be found in section 2.2.1 of Chapter 2. In order to calculate these perfusion measures it is assumed that the contrast agent remains within the vasculature (33). To produce quantitative perfusion measures, an arterial input function (AIF), which is the signal-time course from an artery supplying the tissue of interest, is required (34). An alternative approach is to report relative CBV (rCBV), thereby avoiding the need to measure an AIF, which can be difficult as AIFs can be distorted by partial volume effects (PVEs), affecting the accuracy of the quantitative measures (35, 36). rCBV measurements are calculated by normalising the CBV values from the area of interest to a region of interest (ROI) from another tissue (34). Often this is normally appearing white matter, although any normally appearing tissue can be used (34, 37).

DSC-MRI has applications in indicating health in a range of diseases. For example, it can be used in the diagnosis and treatment of brain tumours. It can be used in the grading of brain tumours, which is important as it determines the treatment received by the patient. It allows a non-invasive diagnosis of the brain tumour prior to confirmation by biopsy, which means that potential treatment can be discussed with the patient and their family at an earlier stage (38, 39). Higher grade tumours are more vascular, which leads to increased perfusion and is reflected in the rCBV values (38). Work from Law et al., Maia et el. and Knopp et al. has shown that rCBV values from gliomas correlate with their grade (38, 40-42). Work by Grist et al. has shown that DSC-MRI rCBV values can be used in combination with metrics extracted from diffusion weighted imaging to distinguish brain tumour type (39). It can be used to identify molecular characteristics of different gliomas (43). Work from Kickingereder et al. showed that tumour rCBV could be used to predict the presence of IDH mutation with those tumours having larger rCBV values (43), and work from Lee et al. finding increased normalised CBV (nCBV), in tumours with the IDH mutation (44). It can be used to differentiate between different types of brain tumour, which is important as different tumours require different treatment pathways (45). Tumour types show variation in their vascularity and therefore differences in blood volume and flow (45). Work by Law et al. has shown that rCBV values from DSC-MRI in combination with magnetic resonance spectroscopy (MRS) can be used to distinguish high grade gliomas (HGG) from metastases (45). Differences in the vasculature between tumour types also lead to differences in blood vessel ‘leakiness’, and work by Cha et al. has used PSR values from DSC-MRI as a measure to distinguish between glioblastoma multiforme (GBM) and metastases (46). It can also be used to plan surgical interventions in brain tumour patients, with work by Barajas et al. showing that rCBV values from DSC-MRI can be used to identify tissue regions surrounding a GBM tumour which need

further treatment (47), whilst work by Lefranc et al. has shown that rCBV values are useful in improving the accuracy of stereotactic biopsies (48). Aside from brain tumours, it has applications in stroke, where it has been used in combination with diffusion weighted imaging (DWI) to identify perfusion mismatches and identify regions which could be salvaged from infarction (49, 50). It also has applications in Alzheimer's, where a decreased CBV has been measured in Alzheimer's patients (51) and age-dependent decreases in CBF and CBV in grey matter have been measured (52).

DSC-MRI has several advantages: it is the most established and widely used MR perfusion technique, meaning there are lots of analysis packages available to analyse the data; it has fast acquisition times of around 2 minutes (34); it offers a better contrast-to-noise ratio than DCE-MRI (53); and it offers better SNR than ASL (53). However, the need to inject contrast agent makes for a more invasive procedure. Gadolinium contrast agents can potentially cause nephrogenic systemic fibrosis in patients with impaired kidney function (54), and there are concerns about potential gadolinium deposition within the brain (55). This is of particular concern in paediatrics as previous individual case studies have shown evidence of gadolinium deposition following repeated doses of contrast agent (56, 57). Therefore, recommended protocols for paediatric DSC-MRI tend to limit the amount of gadolinium which is injected (58). The measurement of perfusion parameters assumes that the blood brain barrier (BBB) is intact and no contrast agent leaks into the extravascular extracellular space (EES), which is not always the case in brain tumour patients (59). In order to fully quantify the perfusion parameters from DSC-MRI an AIF is needed, which can be challenging to obtain (60). DSC-MRI is affected by noise, motion, and susceptibility artefacts, which all impact the quality of the acquired data, and therefore the accuracy of any perfusion parameters estimated from the data (34). The current recommended method for assessing data quality is through qualitative

review, which is very time-consuming and subjective between reviewers (34). These factors, plus variability in acquisition protocol applied between centres, limit the clinical applicability of the technique (61).

### 1.2.2. DCE-MRI

In DCE-MRI, the passage of a gadolinium based contrast agent is imaged using a $T_1$ weighted imaging sequence (62). The contrast agent causes a signal increase due to $T_1$ shortening as it passes through and the change in intensity with time can be plotted as a time intensity curve for each pixel (62). The shape of each time intensity curve depends on the vascular properties of the tissue, and fitting a pharmacokinetic model to the concentration-time curves allows for estimates of physiological parameters such as the volume transfer constant, $K^{trans}$, the volume of extravascular extracellular space per unit volume of tissue, $V_e$, the plasma volume, $V_p$, and the rate constant between EES and blood plasma, $k_{ep}^{**}$, to be calculated (63, 64). It is also possible to determine estimates of the tissue plasma perfusion, $F_p$, and the capillary permeability-surface area product (PS) (65).

DCE-MRI has several applications in relation to brain tumours, where it has been used to distinguish gliomas from radiation injury (66), determine tumour grade (67), predict survival (68), and determine progression from pseudo-progression (69). It has also been used to assess tumours throughout the body, including in the lungs (70), breast (71) and prostate (72). It has applications in assessing kidney function, where it is used to assess the filtration of the kidneys (73, 74). Finally, it has been used to assess atherosclerotic plaques, where $K^{trans}$ has been shown to be a biomarker of plaque inflammation (75).

One of the main advantages of DCE-MRI is its ability to assess vessel leakiness (53), which is very useful in assessing brain tumour grade (76). DCE-MRI suffers from some of the same issues as DSC-MRI as it requires high temporal-resolution scanning, the injection of a

gadolinium based contrast agent (61), and an AIF is needed to estimate the physiological parameters (77). It has a longer acquisition time than DSC-MRI as sufficient leakage needs to have occurred for leakage parameters, such as $V_e$, to be estimated (23). Its clinical utility is also limited by the fact that there are multiple acquisition methods and analysis models available (63).

### 1.2.3. ASL

In ASL, magnetically labelled arterial blood is used to provide estimates of CBF within the brain (26). The arterial blood is labelled by inverting the magnetisation of its protons using a radiofrequency (RF) pulse (78). The arterial blood is labelled outside of the region to be imaged, usually labelling occurs in an artery supplying the region of the body to be imaged. For example, in brain imaging one of the arteries in the neck is labelled (79). A post labelling delay is applied, which is a period of time during which no RF pulses are applied, and no images are acquired (78). This delay allows the labelled protons in the arterial blood to diffuse into the tissue of the region to be imaged (26, 78). An image is then acquired and is referred to as the labelled image (78). A control image is then acquired by repeating the process without any labelling (78). Subtraction of the control image from the labelled image produces a signal difference map which can be used to provide estimates of CBF (26).

ASL has several applications in brain tumours, where it can be used to characterise tumour type (80, 81), classify tumour grade (82, 83) and differentiate between progression and pseudoprogression (84). It also has several applications in age-related disease, where it has been used to identify Alzheimer's disease and has shown comparable results to positron emission tomography (PET) (85, 86). Finally it has also been used to monitor disease progression in patients with Parkinson's disease (87).

One of the main advantages of ASL is that it is non-invasive as it uses an endogenous contrast agent (78). This makes it well suited for patients with reduced kidney function, for use in studies where repeated measurements are required, for use in healthy volunteers, and in paediatrics (88). ASL has been shown to be comparable to DSC, both in terms of the quality of perfusion maps produced by both methods (89), and in terms of significant correlations in the CBF values produced (90, 91). However, ASL has low SNR compared to the other perfusion MR methods (78). There is a delay between labelling and acquisition, which can lead to systematic measurement errors and also means the acquisition time is longer than DSC-MRI and DCE-MRI (88). The choice of post-labelling delay has an effect on the quality of the perfusion measurements and selecting the optimal post-labelling delay in paediatrics and disease states can be challenging (26, 92).

### 1.2.4. IVIM

IVIM uses diffusion-weighted MRI to provide estimates of perfusion within the body, therefore no contrast agents are required (27). In diffusion MRI, diffusion sensitising gradients are applied during the acquisition, the strength and duration of which are determined by the b-values (93). A larger b-value results in the application of the diffusion gradient for a longer time period, leading to a larger diffusion effect (94). The signal after the diffusion sensitive gradients have been applied, S, is dependent on the baseline signal $S_0$, the b-value, b, and the apparent diffusion coefficient (ADC), D, as shown in Equation 1.3 (94).

$$\frac{S}{S_0} = e^{-bD} \tag{1.3}$$

The b-value is selected by the user prior to acquisition and is used to determine how much diffusion-weighting is present in the acquired image (94). The ADC is a measure of the amount of diffusion present (94). Several diffusion weighted images are acquired over a range

of b-values (95). For each pixel, the b-values are plotted against signal, and equation 1.3 can be fitted to the resulting curve to produce an estimate of the ADC values. The bi-exponential IVIM model, shown in equation 1.4, can also be fitted to the b-value curve (96). This model assumes that the b-value curve is the sum of diffusion from tissue and blood components and allows estimates of the flowing vascular volume fraction, f, and the pseudo-diffusion coefficient, $D^*$, to be calculated (96).

$$\frac{S}{S_0} = fe^{-bD^*} + (1-f)e^{-bD} \tag{1.4}$$

Figure 1.2 shows an example of the natural log of the normalised signal ($S/S_0$) plotted against the b-values. This allows both the ADC and IVIM models (represented by equations 1.3 and 1.4, respectively) to be fitted to the data.

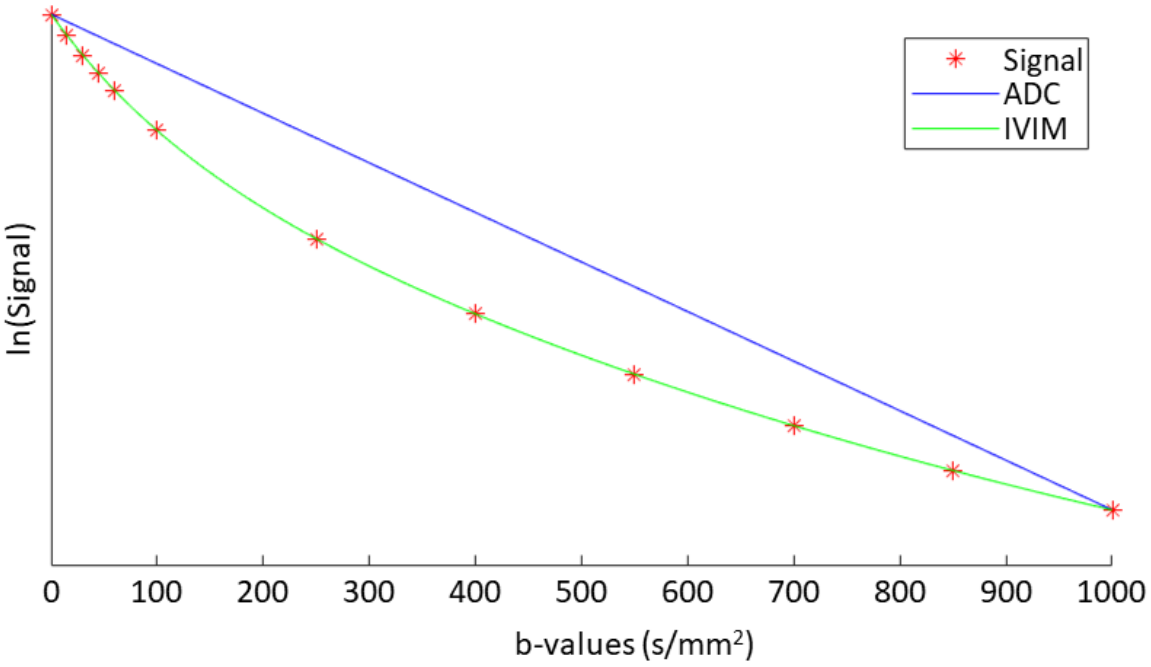



*Figure 1.2: An example of the curve produced from plotting the change in signal with b-value, with the ADC and IVIM models fitted to the curve.*

IVIM has multiple applications in assessing health in different types of disease. For example in brain imaging it has been shown that f moderately correlates with CBV from DSC-MRI

(97), and can be used to determine tumour grade (97, 98). IVIM has applications in stroke where it has been used to identify infarcted regions (99). Finally, IVIM also has applications in body imaging where it has been used to characterise liver lesions (100) and assess kidney function (101).

One of the advantages of IVIM is that it relies on diffusion to estimate blood flow, so does not require any contrast agent, and is completely non-invasive (96). Another advantage is that it is capable of providing a simultaneous measurement of perfusion and diffusion (102). Finally, the signal has high spatial specificity since it originates from the region it is measured in and is independent of the arterial blood flow (103). The main disadvantage of IVIM is that there is no current consensus on the optimal set of b values used in the acquisition, meaning there is a lot of variability between studies (99, 104-106). This limits the usefulness of IVIM as the selection of b values used has a large effect on the perfusion measures obtained from IVIM (107). Finally, there are several different variations on the IVIM model available, with each model having an effect on the perfusion parameters obtained (27).

| Imaging Technique | Parameters measured | Pros | Cons |
|---|---|---|---|
| CT perfusion | CBV, CBF, MTT | - Quick image acquisition<br>- Widely available compared to other methods | - Requires additional radiation dose compared to standard CT<br>- Risks associated with iodinated contrast agents |
| PET | regional CBV, regional CBF and regional CMRO | - High sensitivity<br>- Can provide accurate estimation of tracer concentration<br>- Good diagnostic accuracy | - Limited SNR<br>- Requires a cyclotron to produce the tracer |
| Ultrasound | MTT and TTP | - Very repeatable measurement | - Susceptible to attenuation artefacts |

| | | | |
|---|---|---|---|
| | | - Very portable compared to other methods (can be performed at the bedside) | - Poor sensitivity to slow blood flow and capillary flow |
| NIRS | Oxygen saturation | - Non-invasive imaging | - Limited spatial resolution<br>- Signal can be contaminated by contributions from other molecules |
| DSC-MRI | rCBV, CBV, CBF, MTT | - Most established and widely used MR perfusion technique<br>- Offers the best SNR and contrast to noise ratio<br>- Fast acquisition time | - Requires gadolinium contrast agent<br>- Contrast agent leakage can occur in areas of blood-brain-barrier breakdown<br>- Susceptibility artefacts<br>- Need to normalise to normal tissue unless an AIF is measured |
| DCE-MRI | $K^{trans}$, $V_e$, $V_p$, $k_{ep}^{**}$, $F_p$, PS | - Useful for assessing the leakiness of the BBB, which has multiple applications in assessing brain tumours.<br>- Use in solid body tumours | - Requires the injection of a gadolinium based contrast agent<br>- Clinical utility limited due to multiple acquisition methods and analysis models available<br>- Requires a pre-contrast T1 measurement<br>- Requires measurement of an arterial input function |
| ASL | CBF | - Non-invasive<br>- Well suited for patients with reduced kidney function, for repeated measurements, for use in healthy volunteers, and in paediatrics.<br>- Comparable CBF values to DSC-MRI | - Low SNR compared to contrast agent based methods<br>- Delay between labelling and acquisition can lead to systematic measurement errors<br>- Long acquisition time compared to DSC-MRI and DCE-MRI |
| IVIM | f, D* | - Does not require any contrast agent, and is completely non-invasive | - No current consensus on the ideal b value sequence<br>- Multiple fitting models available which all affect the parameters produced |

| | | | |
|---|---|---|---|
| | | - Provides a simultaneous measurement of perfusion and diffusion<br>- High spatial specificity | |

*Table 1.1: A summary of all the imaging methods discussed, the parameters they measure and the pros and cons of each method.*

## 1.3. Obtaining Robust Biomarkers

A biomarker can be defined as a measurable characteristic that is an indicator of a biological process, pathological change or response to treatment (108). Biomarkers can be obtained from: biofluids, solid tissue samples, physiological measures and imaging measures (109). Imaging biomarkers have the advantage of being relatively non-invasive and inherently quantitative, and also allow for repeated measurements (110). They can be categorised as either a structural, physical, functional, morphological, or textural property (4). They have multiple applications and can be used for: prediction, to determine which patients are more susceptible to a disease; detection, to identify which patients have a disease; staging, to classify the extent of a disease; grading, to indicate disease aggressiveness; prognosis, to determine the likelihood of survival or disease recurrence; and assessment of treatment response, to determine how the disease has responded to a treatment (108).

Often in clinical settings, the interpretation of medical images is carried out by visual assessment, relying on acquired knowledge to identify abnormalities (111). Semi-quantitative approaches, which use categorical scoring systems based on user observations are also widely used; for example, the reporting and data systems (RADS) (112, 113). However, in some cases a quantitative imaging biomarker (QIB) offers a more accurate and objective assessment than manual assessment (111). Widespread use of QIBs in clinical practice is hampered by difficulties in harmonising data acquisition and analysis protocols (111).

Ensuring that QIBs are robust is an important step in validating them for application to a clinical setting (4). The European Imaging Biomarkers Alliance (EIBALL) has produced a set of recommendations for using validated imaging biomarkers in clinical decision making (111). Whilst a collaboration between the Quantitative Imaging Biomarker Alliance (QIBA) and the Radiological Society of North America (RSNA) has produced a statistical framework to assess the technical performance of potential QIBs (4). QIBA has also recently published a set of recommendations to improve precision of biomarkers obtained from DWI and DCE-MRI in multicentre oncology trials (114). An example of some perfusion imaging biomarkers and their applications are shown in Table 1.2.

As can be seen from table 1.2, perfusion biomarkers have many applications in identifying and evaluating disease and may have an impact on a patient's treatment if used clinically. The accuracy of the biomarker values is dependent on the quality of the data acquired, which they have been estimated from. Therefore, determining data quality is an important step in determining biomarker accuracy, and improved data quality can lead to improved biomarker accuracy.

| Imaging Biomarker | Imaging Modality | Applications |
|---|---|---|
| Left ventricular ejection fraction (LVEF) | Contrast-enhanced Echocardiography (Ultrasound) | Assess heart function (115) |
| Blood flow, blood volume and permeability surface product | CT Perfusion | Identifying head and neck squamous cell carcinoma (116) |
| | | Identifying liver cancer (117) |
| | | Blood flow useful for evaluating rectal cancer (118) |
| | | Permeability can predict survival from lung cancer (119) |
| $K^{trans}$, $K_{ep}$, blood flow, $V_e$ | DCE-MRI | Identifying recurrent glioblastoma (120) |
| | | Identifying liver cancer (117) |
| | | Assessing treatment response (121) |
| rCBV, CBV, CBF, MTT | DSC-MRI | Discriminating glioma grade (122) |
| | | Differentiating tumour progression from treatment effects (123, 124) |

| | | |
|---|---|---|
| | | Identifying perfusion delay in patients with Moyamoya disease (125) |
| | | Identifying regions of the brain which would benefit from thrombolytic therapy following stroke (49) |

*Table 1.2: Some examples of perfusion imaging biomarkers and their applications.*

## 1.4. Aims and Objectives

The perfusion parameters obtained from DSC-MRI (CBF, CBV and MTT) have applications as biomarkers for disease. In order to obtain robust biomarkers, the perfusion parameters must be accurate. Since estimates of the perfusion parameters are derived from the acquired data, ensuring that the data acquired is of sufficient quality to obtain accurate perfusion estimates is crucial in obtaining robust biomarkers. Therefore, the overall aim for this project was to investigate methods for improved processing of DSC-MRI data, validate them, and then combine them to produce an automated processing pipeline for DSC-MRI, which could assess data quality and produce robust biomarkers.

The objectives of this project were to:

1. Assess the performance of the recommended processes for assessing DSC-MRI data quality. Investigate ways to replace this with an automated method using conventional methods and machine learning.

2. Investigate the performance of differing denoising methods to improve the quality of DSC-MRI data. Develop a DSC-MRI simulator to validate each denoising method and quantify the differences in performance between methods. The simulator should be capable of modelling and accommodating for a range of acquisition protocols.

3. Develop an automated pipeline capable of segmenting the relevant regions of the brain, denoising the data, assessing data quality and producing estimates of perfusion parameters.

4. Apply the automated pipeline to a set of patient data.

## 1.5. Thesis Structure

Chapter 2 presents the theory of MRI and the computational techniques used. In the MRI section, an overview of the physics behind MRI is presented, along with an explanation of how DSC-MRI uses a contrast agent to provide estimates of perfusion. Common acquisition protocols and analysis methods for DSC-MRI are covered. In the computational techniques section, the machine learning, denoising methods and statistical tests applied throughout this piece of work are introduced.

Chapter 3 reviews the performance of qualitative review as a method for determining data quality and presents the development of an automated alternative. This uses parameters extracted from the DSC-MRI signal-time courses and investigates whether machine learning can be used to predict data quality.

Chapter 4 looks into potential denoising methods to improve the quality of DSC-MRI data. This focuses on the application of matricization, wavelets and the Tucker decomposition as denoising methods. All three denoising methods are applied to simulated and patient data and their performance is assessed.

Chapter 5 presents the results from combining the machine learning from Chapter 3 and the denoising from Chapter 4 to create an automated processing pipeline for DSC-MRI. Also presented are the results of applying the pipeline to a multicentre dataset.

Chapter 6 presents the results of applying the automated pipeline from Chapter 5 to DSC-MRI data from a single centre study of brain tumour survivors who received differing treatments.

Chapter 7 draws conclusions from the work presented in Chapters 3-6 and discusses possible future work.

# Chapter 2: Theory

# 2. Theory

## 2.1. MRI

### 2.1.1. Nuclear Magnetic Resonance (NMR)

MRI relies on the concept of nuclear magnetic resonance (NMR) to produce images, which in turn relies on nuclei with magnetic properties. A nucleus has magnetic properties when it precesses about its nuclear axis, generating a magnetic moment. The magnetic moment, μ, is related to the angular momentum of the nucleus, P, through equation 2.1, where γ is the gyromagnetic ratio. P is related to the spin quantum number of the nucleus, I, through equation 2.2, where h is Planck's constant (126).

$$\mu = \gamma P \tag{2.1}$$

$$P = \sqrt{I(I+1)}\,\frac{h}{2\pi} \tag{2.2}$$

Any nucleus with a non-zero value for I will have magnetic properties. The value of I for a given nucleus is determined by the number of unpaired nucleons. This is because nucleons form proton-proton and neutron-neutron pairs cancelling out their angular momentum. Therefore, a nucleus with an even number of protons and neutrons will have no unpaired nucleons and I = 0. All nucleons have I = ½, so a nucleus with an odd number of unpaired protons or neutrons will have a half integer value for I, whilst a nucleus with an even number of unpaired protons and neutrons will have an integer value for I (126).

Some examples of nuclei which have non-zero values for I, and consequently have magnetic properties, include $^{1}$H, $^{23}$Na, $^{31}$P and $^{129}$Xe. $^{1}$H, which consists of a single proton, is the most abundant element within the body and is therefore used in MR imaging (127). When there is no magnetic field present, all the magnetic moments of $^{1}$H nuclei are randomly aligned. As each $^{1}$H nuclei has a magnetic moment of equal magnitude, the vector sum of all the magnetic

moments will add up to zero, and there will be no overall net magnetisation (127). When a magnetic field, $B_0$, is applied the $^1H$ nuclei can exist in one of $(2I + 1)$ quantised states. The $B_0$ field is applied along the z axis, so the angular momentum, $P_z$, of each $^1H$ nuclei will exist in one of $(2I + 1)$ orientations respective of the $B_0$ field. Hence there is also a magnetic moment, $\mu_z$, and energy, E associated with each quantised state. Equation 2.3 summarises the relationship between the energy, magnetic moment, and angular momentum, where m has $(2I + 1)$ values $m = I, I - 1, I - 2, \dots, -I$.

$$E = \mu_z B_0 = \gamma P_z B_0 = \gamma m \hbar B_0 \tag{2.3}$$

For $^1H$, I = ½, so there are two possible states: the spin-up state where the $^1H$ nucleus is aligned to $B_0$; or the spin-down state where the $^1H$ nucleus is anti-aligned to $B_0$. Figure 2.1 summarises the two possible orientations, their associated energy, and the difference in energy between the two states, ΔE.

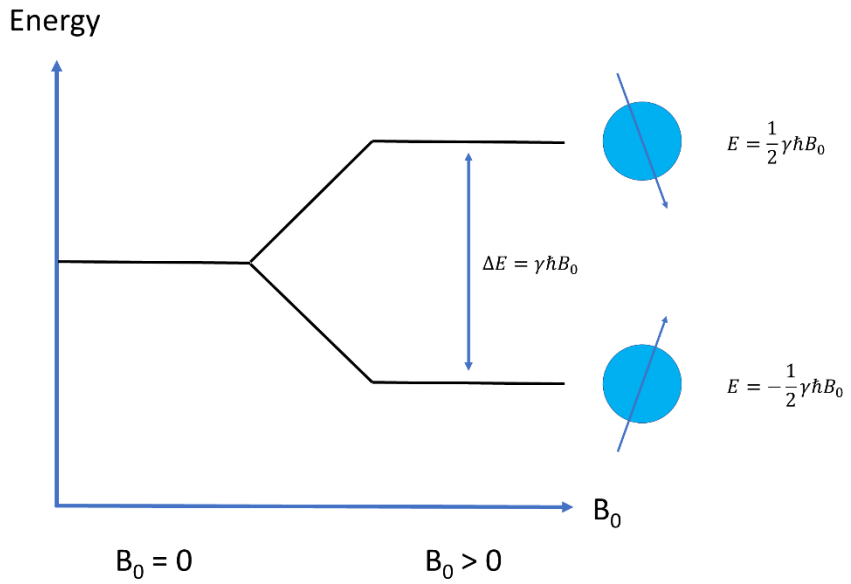



*Figure 2.1: The possible orientations of the $^1H$ nuclei in a magnetic field.*

Due to the quantised nature of the two states, neither the spin-up or spin-down energy states are perfectly aligned to the magnetic field. This means that the magnetic moment is oriented at an angle θ to the $B_0$ field which it precesses around, as shown in Figure 2.2. This precession is caused by the force exerted on the $^1H$ nucleus by the $B_0$ field, called a torque, The torque tries to force the $^1H$ nuclei to align with exactly with the field lines. However, as the $^1H$ nuclei can only exist in one of two quantised energy states, which are at an angle to the $B_0$ field, it experiences a constant torque, causing it to precess about the $B_0$ field (128).

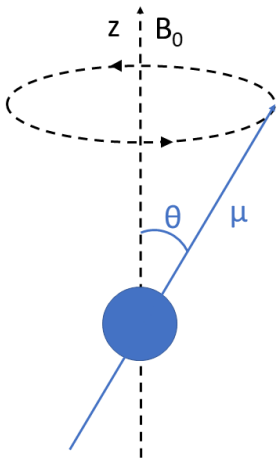



*Figure 2.2: A magnetic moment precessing around the $B_0$ field.*

The magnetic moment precesses at the Larmor frequency, $\omega_L$, which is related to the $B_0$ field and is given by equation 2.4 (128). $^1H$ has a gyromagnetic ratio of 42.58 MHz $T^{-1}$, which produces $\omega_L$ values of 63.87 MHz and 127.74 MHz at field strengths of 1.5 T and 3 T, respectively (129).

$$\omega_L = \gamma B_0 \tag{2.4}$$

Vector addition of all the magnetic moments now produces a different result compared to when there was no $B_0$ field applied. Perpendicular to the $B_0$ field (in the xy plane) there is no net magnetisation. This is because the precessions of the individual magnetic moments are all randomly oriented, so cancel each other out. Parallel to the $B_0$ field (along the z axis) there is a

net magnetisation, $M_0$. This is because the spin-up state is lower energy than the spin-down state, so slightly more magnetic moments will be in the spin-up state. The orientation of a proton depends on its energy and the distribution of protons between the spin-up and spin-down state is defined by the Boltzmann distribution and is shown in equation 2.5. $N_{up}$ and $N_{down}$ are the numbers of protons in the spin-up and spin-down states respectively, $\Delta E$ is the energy difference between the spin-up and spin-down states, $k_b$ is the Boltzmann constant and T is the temperature (128).

$$\frac{N_{up}}{N_{down}} = e^{\frac{\Delta E}{k_B T}} \tag{2.5}$$

The value of $M_0$ depends on the strength of the $B_0$ field applied and is given by equation 2.6, where $\chi$ is the magnetic susceptibility of the substance (127).

$$M_0 = \chi B_0 \tag{2.6}$$

### 2.1.2. Radiofrequency (RF) Pulses

$M_0$ is very small compared to $B_0$, so it is not possible to measure the net magnetisation whilst it is parallel to the $B_0$ field. It is possible to tip the net magnetisation into the xy plane by applying an oscillating $B_1$ field, in the form of a radiofrequency (RF) pulse. In order to have an effect on $M_0$, by inducing resonance, the RF pulse must have a frequency equal to $\omega_L$. These frequencies are necessary because the RF pulse must have an energy equal to $\Delta E$, the energy gap between the two levels. Substituting, the Larmor equation (2.4) into the equation for $\Delta E$ (Figure 2.1) gives equation 2.7.

$$\Delta E = \hbar \omega_L \tag{2.7}$$

The angle between $M_0$ and $B_0$ following the application of the RF pulse is known as the flip angle (FA), α. The FA produced by RF pulse depends on the strength of its $B_1$ field and the length of time the pulse is applied for, $t_p$, and is given by equation 2.8 (128).

$$\alpha = \gamma B_1 t_p \tag{2.8}$$

As well as tipping $M_0$ by an angle α, the RF pulse also has the effect of synchronising the precession of all the individual magnetic moments, bringing them to phase coherence. This means that following the application of the RF pulse there is a net $M_0$ precessing around the $B_0$ field with a component perpendicular to the $B_0$ field. The transverse component, $M_{xy}$, depends on the flip angle of the RF pulse applied. Figure 2.3 shows an example of an RF pulse being applied to the net magnetisation and the resulting changes to the transverse, $M_{xy}$, and longitudinal, $M_z$, components of the magnetisation. This diagram uses the rotating reference frame, where the xy plane rotates about the z axis at $\omega_L$. In this reference frame $M_0$ can be considered to be stationary (128).

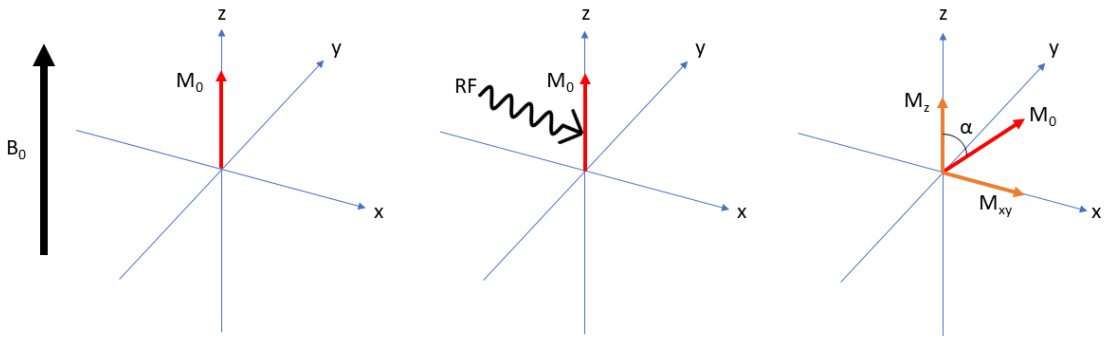



*Figure 2.3: The rotation of the net magnetisation following the application of an RF pulse.*

A commonly used RF pulse in MR imaging is the 90˚ pulse, which tips the entirety of $M_0$ into the transverse place, resulting in a maximum value for $M_{xy}$. The precessions of the $M_{xy}$

component induce an oscillating voltage in the receiver coil of the MRI scanner, which takes the form of a sine wave (128).

### 2.1.3. $T_1$ and $T_2$ Relaxation

Once the RF pulse is turned off the individual $^1H$ nuclei undergo relaxation, where they begin to realign with the field lines of the $B_0$ field, by reverting to their spin-up and spin-down states, and their precessions begin to de-phase. Both of these relaxation processes cause the magnitude of the transverse component of $M_0$ to decay. This produces a free induction decay (FID) signal in the receiver coil, which is a sine wave damped by an exponential and is shown in Figure 2.4 (128).

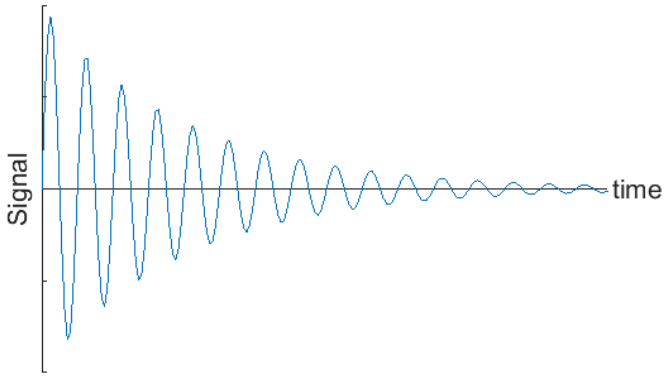



*Figure 2.4: The free induction decay (FID) signal.*

$T_1$ relaxation, also referred to as spin-lattice relaxation, describes the mechanism by which protons release the energy they absorbed from the RF pulse and realign with the $B_o$ field lines. It is referred to as spin-lattice relaxation as it describes the interactions between the anti-aligned $^1H$ nuclei and surrounding molecules, which results in the $^1H$ nuclei transferring their energy and re-aligning with the $B_0$ field. The $T_1$ relaxation time is the time it takes for longitudinal component of $M_0$ (parallel to $B_0$) to return to 63% of its original value, as shown in Figure 2.5. The change in longitudinal magnetisation, $M_z$, as the $^1H$ nuclei undergo $T_1$ relaxation is shown in equation 2.9, where $T_1$ is the $T_1$ relaxation time and t is the time since

the RF pulse was turned off (127). Figure 2.6 demonstrates how the magnitude of the $M_z$ vector changes with time, from the rotating reference frame, following the application of a 90° RF pulse.

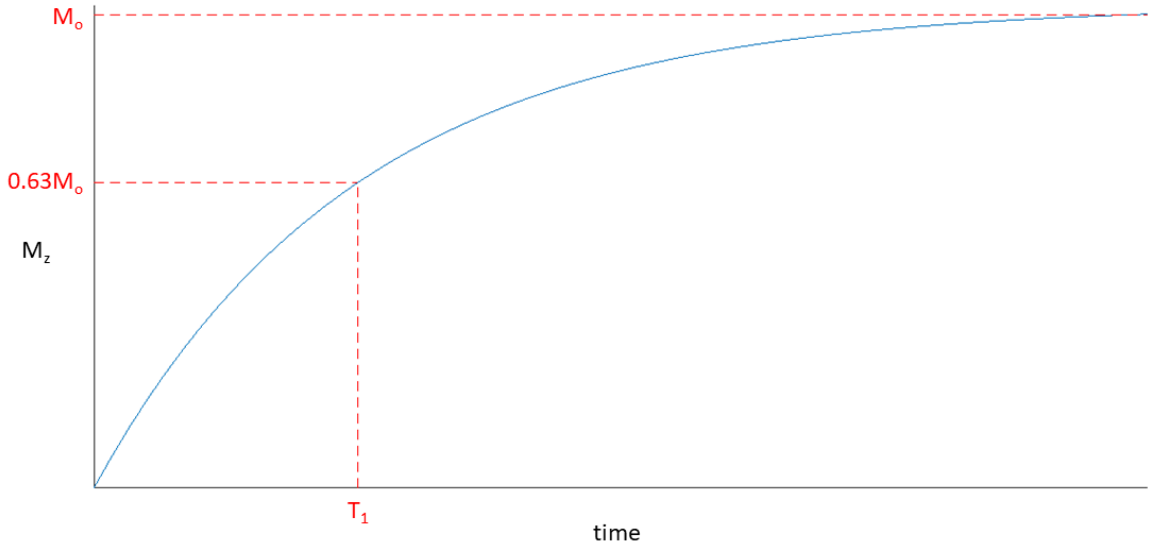



*Figure 2.5: Change in longitudinal magnetisation with time after the RF pulse is switched off.*

$$M_z(t) = M_0\left(1 - e^{-\frac{t}{T_1}}\right) \tag{2.9}$$

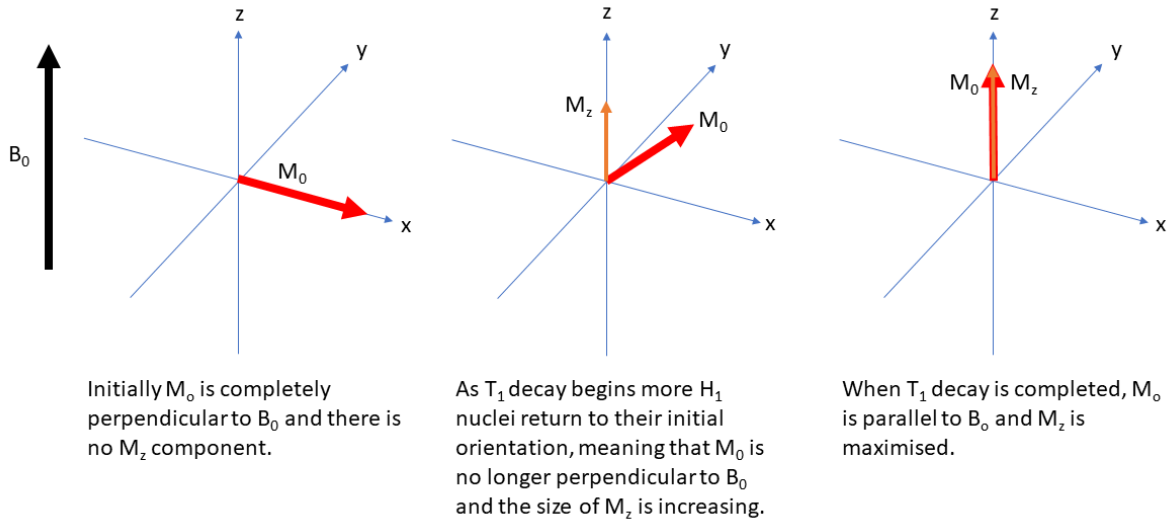



*Figure 2.6: The change in $M_z$ with time during the $T_1$ relaxation process following a 90° RF pulse.*

$T_2$ relaxation, also referred to as spin-spin relaxation is the dephasing of the precessing $^1H$ nuclei after the RF pulse is turned off. It is referred to as spin-spin relaxation because it is

influenced by the interactions between neighbouring $^{1}H$ nuclei. The movement of $^{1}H$ nuclei close to each other causes forces to develop between the nuclei, which in turn leads to local variations in the magnetic field. This causes local differences in $\omega_L$, which leads to a loss of phase and therefore a dispersal of the transverse magnetisation (perpendicular to the $B_0$ field), $M_{xy}$ (130). The $T_2$ relaxation time is the time constant for the $T_2$ decay and is the time taken for $M_{xy}$ to decay to 37% of its maximum value, as shown in Figure 2.7. The decay of $M_{xy}$ is described by equation 2.10 (127). Figure 2.8 demonstrates how the magnitude of $M_{xy}$ changes with time, from the rotating reference frame, following the application of a 90° RF pulse.

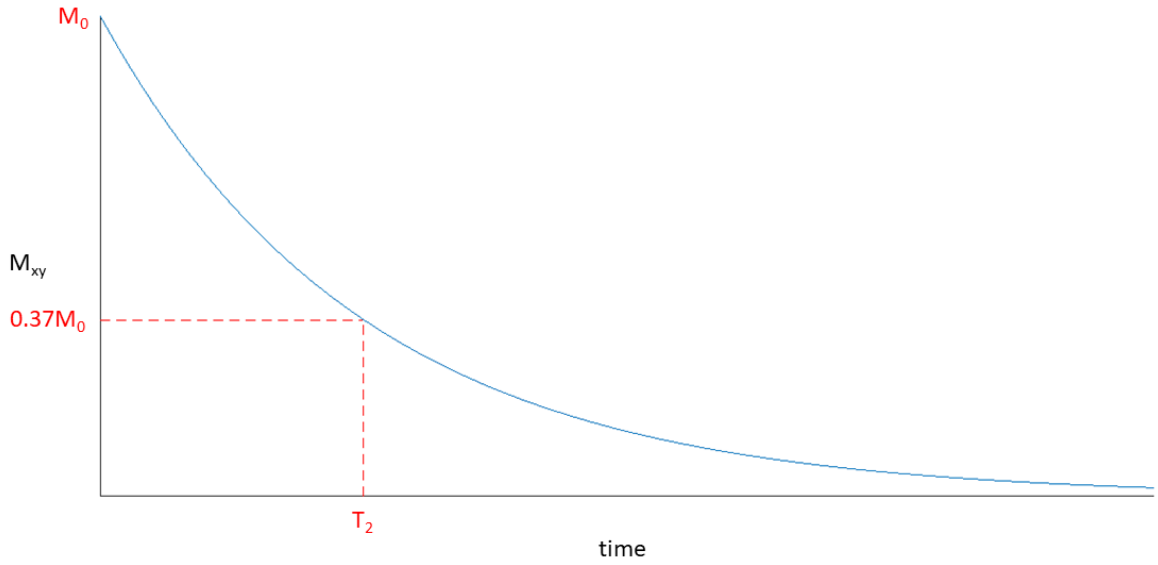
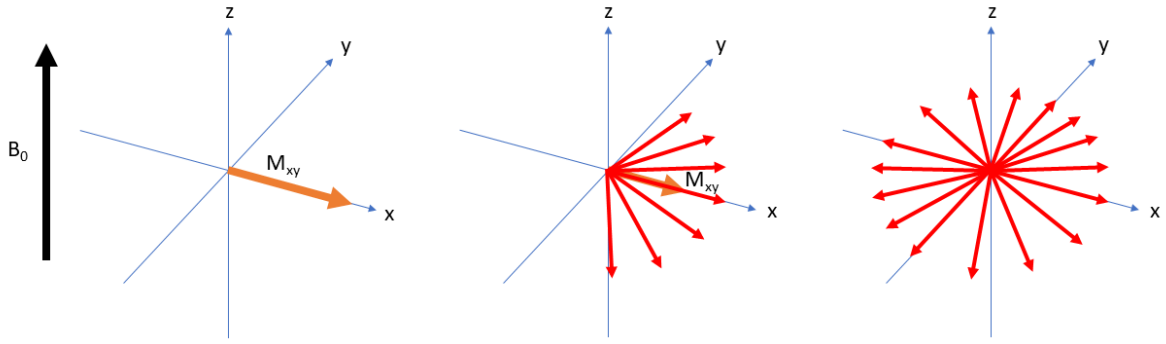



*Figure 2.7: Change in transverse magnetisation with time, after the RF pulse is switched off.*

$$M_{xy}(t) = M_0 e^{-\frac{t}{T_2}} \qquad (2.10)$$




Initially all $^1$H nuclei are in phase and are precessing at the same frequency. $M_{xy}$ is at its maximum value

As $T_2$ relaxation begins the $^1$H nuclei begin to fan out and the magnitude of $M_{xy}$ decreases.

After $T_2$ relaxation is completed all the $^1$H nuclei have completely fanned out reducing the magnitude of $M_{xy}$ to zero.

*Figure 2.8: The change in $M_{xy}$ with time during $T_2$ relaxation following a 90° RF pulse.*

Inhomogeneities in the magnetic field also contribute to spin-spin dephasing. The $T_2^*$ decay time accounts for both tissue specific $T_2$ decay, and decay caused by field inhomogeneities and is given by equation 2.11 (127).

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B_0 \qquad (2.11)$$

### 2.1.4. Magnetic Field Gradients

Magnetic field gradients are an important part of image acquisition in MRI. A magnetic field gradient induces a gradient in the $B_0$ field in the direction it is applied. They are applied over short time intervals during imaging and are small compared to the $B_0$ field, inducing changes of less than 1% (127). Figure 2.9 shows an example of a magnetic field gradient G applied in the x direction and how it affects the magnetic field, $B_x$. The dashed line shows the value of $B_x$ across the x direction when $G = 0$, whilst the solid line shows the value of $B_x$ when $G \neq 0$.

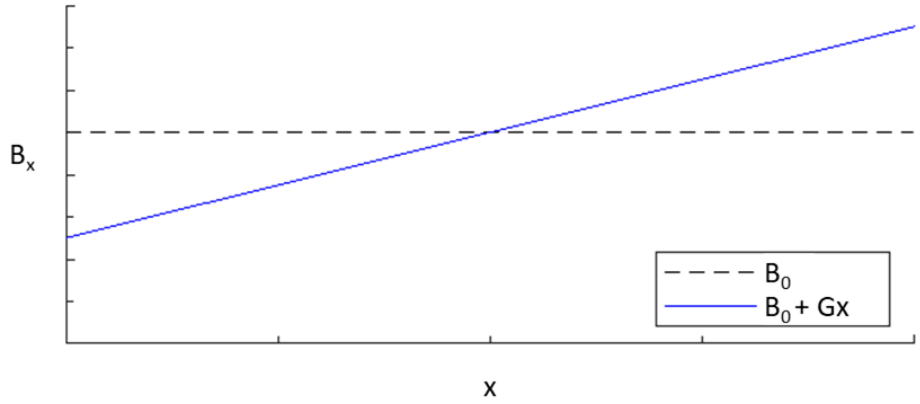



*Figure 2.9: An example of a magnetic field gradient, G, and how it affects the magnetic field in the x direction, $B_x$.*

### 2.1.5. Producing a Signal

Applying a 90˚ RF pulse induces an FID signal which starts to decay as soon as the RF pulse is turned off. In order to produce a recordable signal, it is necessary to refocus the dephasing spins. Rephasing the spins is referred to as producing an echo and can be created by either an RF pulse (spin echo) or a magnetic field gradient (gradient echo).

A spin echo is created by a combination of a 90˚ pulse and a 180˚ pulse. Once the 90˚ pulse has been applied the spins start to dephase. The 180˚ pulse flips all the spins by 180˚. As all the spins are still precessing at the same frequency they are now rephasing rather than dephasing. Rephasing occurs until all the spins are in phase and the signal begins to dephase again. This process of rephasing and dephasing is called an echo and the centre of the echo is the point where all the spins are back in phase. Figure 2.10 summarises this process from the rotating reference frame.

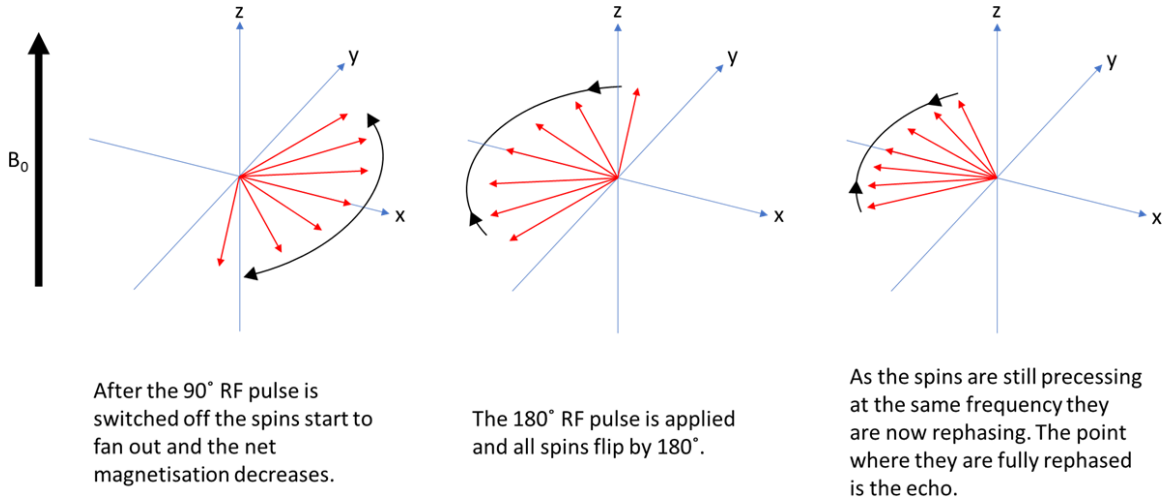



*Figure 2.10: How a spin echo is formed in the rotating frame using a combination of 90 and 180 degree RF pulses.*

A gradient echo is created by replacing the 180˚ RF pulse with a set of gradients. First a dephasing gradient is applied, which accelerates the dephasing of the all the spins, by altering the local magnetic fields and therefore the resonant frequencies. Following this a rephasing gradient is applied, which has the same strength but opposite polarity of the dephasing gradient. The rephasing gradient is applied for double the length of the dephasing gradient. This is because the spins will rephase at the point where the rephasing gradient is of equal area to the dephasing gradient. Therefore, for the echo to be at the centre of the rephasing gradient it needs to be double the area of the dephasing gradient (131).

### 2.1.6 Localising the Signal

To create an image, it is necessary to isolate groups of spins, which is achieved using magnetic field gradients. The first step in localising a signal is the application of a slice select gradient, which induces a gradient in the $B_0$ field along the direction the gradient is applied. The direction of the field gradient defines the orientation of the slice, with the slices being perpendicular to the applied gradient. For example, for an axial image, the slice select gradient is applied in the z direction, parallel to the $B_0$ field lines. Therefore, from equation 2.4. there will be a variation in the Larmor frequencies along the z direction. This means that

a specific slice location and thickness can be selected by using a frequency selective RF pulse containing a range of frequencies. The central frequency of the RF pulse determines the central location of the slice, whilst the range of frequencies controls the thickness of the slice. The range of frequencies is referred to as bandwidth, ΔF, and is related to the slice thickness, Δz and the slice select gradient, $G_{SS}$, through equation 2.12 (127).

$$\Delta F = \gamma G_{SS} \Delta z \tag{2.12}$$

The frequency encoding (or readout) gradient defines one image dimension and is applied perpendicular to the slice select gradient. For an axial image this could be in either the x or y direction. It is applied continuously whilst the signal is recorded by sampling it over a series of equally spaced time points. The gradient causes the protons within the imaging slice to precess with differing frequencies, determined by the location of the proton within the gradient field and the strength of the applied gradient. As the protons are all precessing at differing frequencies, the signal recorded at each time point will therefore correspond to a different spatial frequency (128).

The phase encoding gradient defines the other image dimension and is applied perpendicular to the other two gradients. The effect of applying a phase encoding gradient is demonstrated in Figure 2.11, by considering a phase encoding gradient applied across a column of protons, from the rotating reference frame. Before the phase encoding gradient is applied all protons precess in phase at the same frequency (first column of Figure 2.11). When the phase encoding gradient is switched on, it causes a local change in the $B_0$ field, which leads to changes in the frequency at which the protons precess. In Figure 2.11, the proton in the middle row precesses at the same frequency as it did prior to the application of the phase encoding gradient, as the gradient is zero at this point. The proton in the top row precesses

slower as it experiences a negative phase encoding gradient, and the proton in the bottom row precesses faster as it experiences a positive phase encoding gradient. When the phase encoding gradient is switched off all the protons return to precessing at their original frequency. However, they are now out of phase due to the phase encoding gradient. The difference in phase between the protons depends on the strength of the phase encoding gradient and the length of time it is applied for. In order to produce a complete image, the process of applying gradients is repeated multiple times with phase encoding gradients of differing amplitudes. This is necessary to ensure that enough spatial frequencies are recorded to fully re-create an image. Repeated application of gradients increases the total acquisition time for the scan (127).

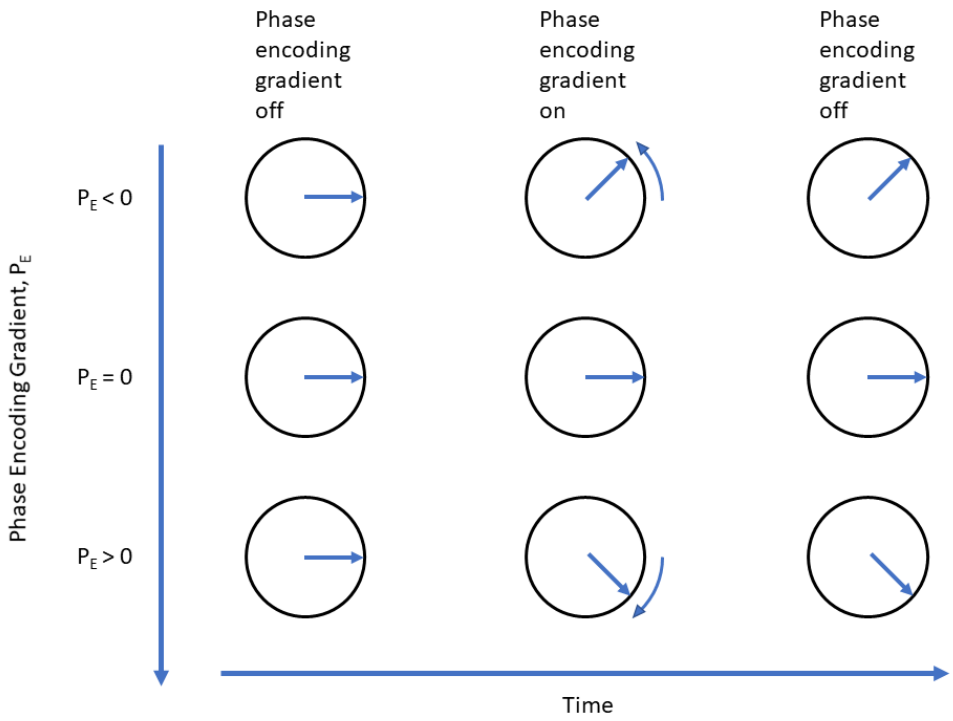



*Figure 2.11: The effect on the individual spins by applying a phase encoding gradient.*

The final acquired signal is a superposition of all the differing precession frequencies, which can be extracted using a Fourier transform. Following the Fourier transform the signal can be

plotted as frequency against amplitude of each frequency, with frequency defining the position along the frequency encoding gradient, and the amplitude of each frequency defining the amount of signal at that position. The Fourier transform of the signal acquired from one frequency encoding and one phase encoding gradient represents a one dimensional projection of the object being imaged (128).

### 2.1.7. Producing an Image

To form an image, RF pulses are combined with magnetic field gradients to create a pulse sequence. The resulting signals detected in the receiver coil are digitised and then stored in k-space, which represents the spatial frequencies of the image. The k-space data is later converted to an image using a Fourier transform. Figure 2.12 shows an example of a spin echo pulse sequence, whilst Figure 2.13 shows an example of a gradient echo pulse sequence. The time between the RF pulse and the centre of the echo is defined as the echo time (TE), whilst the time between two successive excitation pulses is defined as the repetition time (TR) (131).

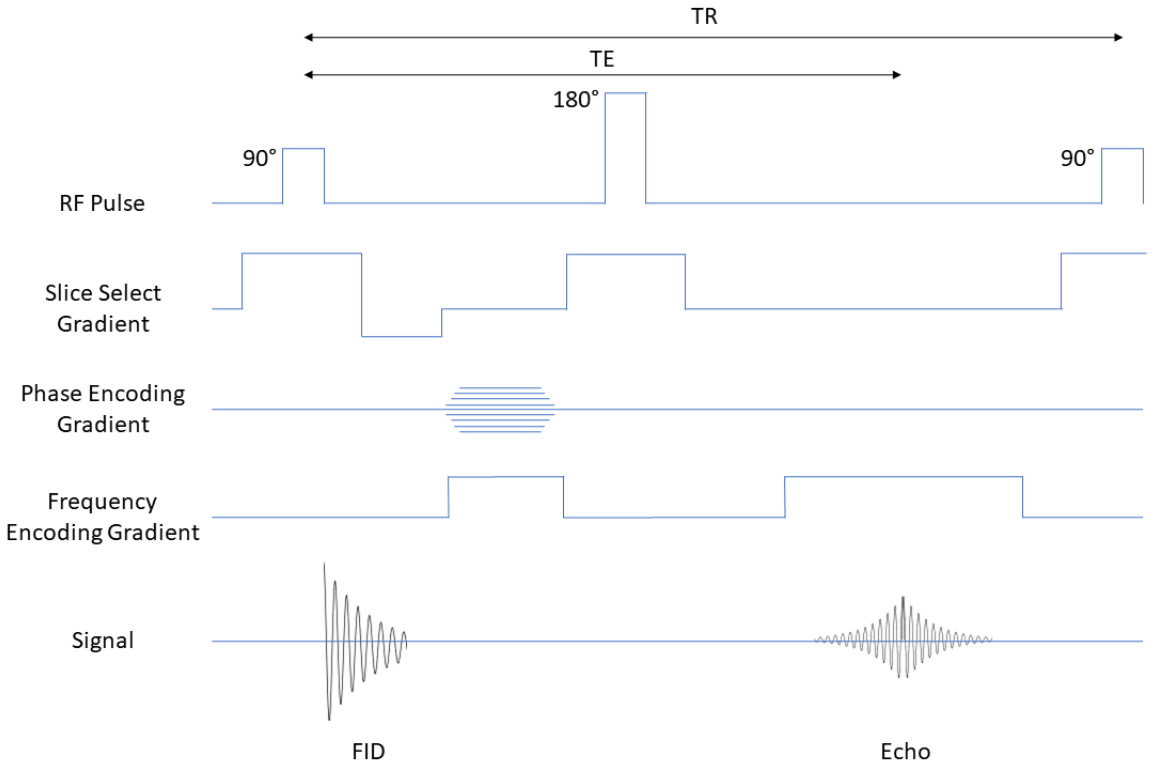



*Figure 2.12: Spin echo pulse sequence diagram.*

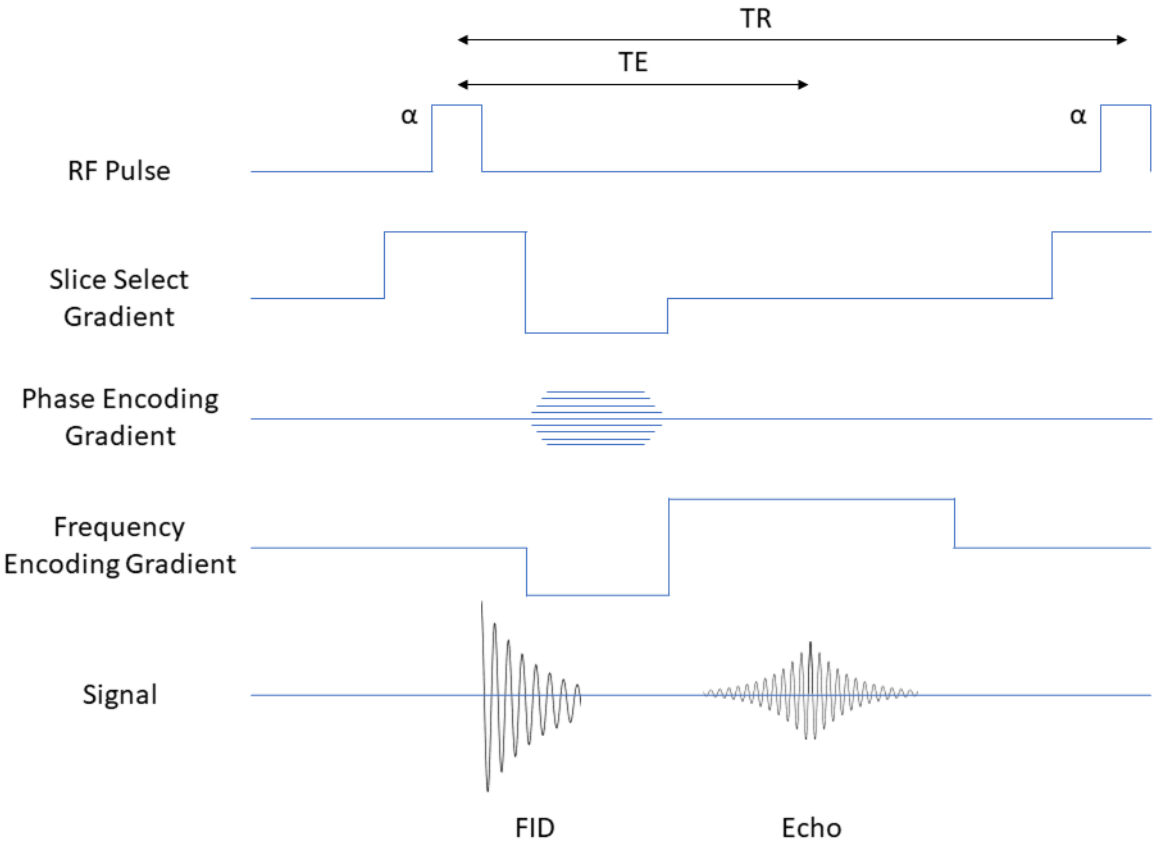



*Figure 2.13: Gradient echo pulse sequence diagram.*

For both pulse sequences, the slice select gradient is applied at the same time as the RF excitation pulse. It is followed by a negative gradient lobe, the purpose of which is to compensate for any dephasing caused by the slice select gradient (128).

For the spin echo pulse sequence, this is followed by the simultaneous application of a frequency and phase encoding gradient. Following this a 180˚ RF pulse, which is used to create the spin echo, is applied at the same time as a second slice select gradient. A second frequency encoding is applied for the duration of the spin echo and a signal is recorded in the receiver coil (128).

For the gradient echo sequence, the slice select gradient is followed by a frequency encoding gradient, which is used to create the gradient echo. The frequency encoding gradient consists of a dephasing lobe and a rephasing lobe. During the dephasing lobe the phase encoding gradient is also applied, and during the rephasing lobe a signal is recorded in the receiver coil (128).

The flip angle of the RF excitation pulse used differs between spin echo and gradient echo pulse sequences. In spin echo the RF excitation pulses usually have flip angles of 90° and 180°, whilst in gradient echo it has a flip angle, α, which is often less than 90°. This is because gradient echo sequences have shorter TR values, which means they have shorter acquisition times than spin echo sequences. The flip angle is therefore smaller to prevent signal saturation from occurring and to make use of the shorter TR (132).

For the pulse sequences above, an entire line of k-space is acquired during the signal acquisition. In order to create a full image multiple lines of k-space must be acquired. Therefore, both pulse sequences must be repeated many times in order to acquire a full image. During each repetition, the amplitude of the phase encoding gradient is changed, and this

defines the $k_y$ value of the line of k-space that is acquired during the rephasing lobe. Once enough lines of k-space are acquired a 2D Fourier transform is applied to the k-space values and an image is produced. The field of view in the phase encoding direction, $FOV_{PE}$, defines the number of phase encoding gradients needed, which in turn defines the image resolution in the phase encoding direction, $\Delta y$. This is because the spacing between k-space lines in the phase encoding direction, $\Delta k_y$, is related to the field of view (FOV) in the phase encoding direction of the acquired image, $FOV_{PE}$, by equation 2.13 (131).

$$\Delta k_y = \frac{1}{FOV_{PE}} \tag{2.13}$$

Multiplying equation 2.13 by the number of phase encoding gradients, $N_{PE}$, gives the $k_y$ value of the largest amplitude phase encoding gradient, $k_{y,max}$, and produces equation 2.14 (131).

$$k_{y,max} = N_{PE}\Delta k_y = \frac{N_{PE}}{FOV_{PE}} = \frac{1}{\Delta y} \tag{2.14}$$

A similar equation can be produced relating the field of view, $FOV_{FE}$, and resolution, $\Delta x$, in the frequency encoding direction to the number of samples recorded during the application of the frequency encoding gradient, $N_{FE}$. The resolution of the final image can be improved by increasing the number of phase encoding gradients and frequency encoding steps over the same field of view. However, increasing the number of phase encoding steps has a large impact on the acquisition time, as shown in equation 2.15, where $T_{Acq}$ is the acquisition time, and NSA is the number of signal averages (128). This means that there is a trade-off between optimising the resolution and acquisition. This is a particular problem for the pulse sequences above, as a single line of k-space is acquired during each TR, meaning that acquisition times are slow (131). Some sequences allow for multiple lines of k-space to be acquired during a

single TR, which reduces acquisition time. Some examples of these will be discussed in section 2.2.2.

$$T_{Acq} = NSA \cdot N_{PE} \cdot TR \quad (2.15)$$

The selection of the $FOV_{PE}$ is particularly important, because if $FOV_{PE}$ is smaller than the anatomy being imaged this leads to a wrap-around artefact where anatomy outside of $FOV_{PE}$ is folded back into the image. This artefact can be prevented by increasing the FOV, but this reduces the resolution in the phase encoding direction. The resolution can be maintained by increasing the number of phase encoding steps, but this increases the scan time. An alternative is to swap the frequency and phase encoding axes, but this only works if $FOV_{PE}$ is larger than the anatomy along its new axis. A third alternative is to use spatial saturation bands to saturate the signal outside of $FOV_{PE}$. However, saturation bands are not capable of fully saturating the signal, so the artefact may not be fully suppressed (128).

### 2.1.8. Weighting of MR Images

The weighting of the acquired MR images is controlled by the acquisition parameters of the pulse sequence used. The parameters used to control this differ between spin echo and gradient echo imaging.

For spin echo imaging the weighting is controlled by the values of TR and TE. $T_1$ weighting is achieved using a short TR and a short TE. The short TR maximises the differences between $T_1$ recovery between tissues. The short TE minimises the amount of dephasing which has occurred, minimising the sensitivity of the acquired signal to changes in $T_2$. Therefore, the acquired signal intensity is dependent on the $T_1$ relaxation rate. $T_2$ weighting is achieved using a long TR and a long TE. The long TR allows for the longitudinal magnetisation to mostly recover between excitation pulses, which minimises the sensitivity of the acquired signal to differences in $T_1$. The long TE allows lots of $T_2$ dephasing to occur. Therefore, the acquired

signal intensity is dependent on the $T_2$ relaxation rate. Proton density (PD) weighting is achieved using a long TR and a short TE, which minimises the effects of both $T_1$ and $T_2$ relaxation on the signal intensity. Therefore, the signal intensity reflects differences in PD between tissues (133).

In gradient echo imaging the TR is always short, which minimises the time allowed for $T_1$ recovery. Therefore, the weighting of gradient echo images is controlled by the flip angle of the excitation pulse and the TE. The amount of $T_1$ weighting is controlled by the flip angle, whilst the amount of $T_2^*$ weighting is controlled by the TE. $T_2^*$ weighting is not possible with spin echo imaging because the 180° pulse eliminates any $T_2^*$ effects (134). $T_1$ weighting is achieved using a large flip angle, short TR and short TE. A larger flip angle means that less of the net $M_0$ remains in the longitudinal direction following excitation, so the acquired signal is more sensitive to the $T_1$ relaxation. Figure 2.14 shows how changing the flip angle changes the size of $M_z$. As with spin echo imaging the short TE minimises the amount of dephasing, which minimises the sensitivity of the acquired signal to $T_2^*$ relaxation. Therefore, the signal is dependent on the $T_1$ relaxation rate. $T_2^*$ weighting is achieved using a low flip angle, short TR and long TE. The low flip angle means that more of the net $M_0$ remains in the longitudinal direction following excitation, so the acquired signal is less sensitive to the $T_1$ relaxation. A long TE maximises the differences in $T_2^*$ recovery between tissues. Therefore, the acquired signal is dependent on the $T_2^*$ relaxation rate. PD weighted images are produced using a low flip angle, short TR and short TE. As with spin echo imaging, this minimises the sensitivity of the acquired signal to the $T_1$ and $T_2^*$ relaxation, meaning the acquired signal is most sensitive to the PD (127).

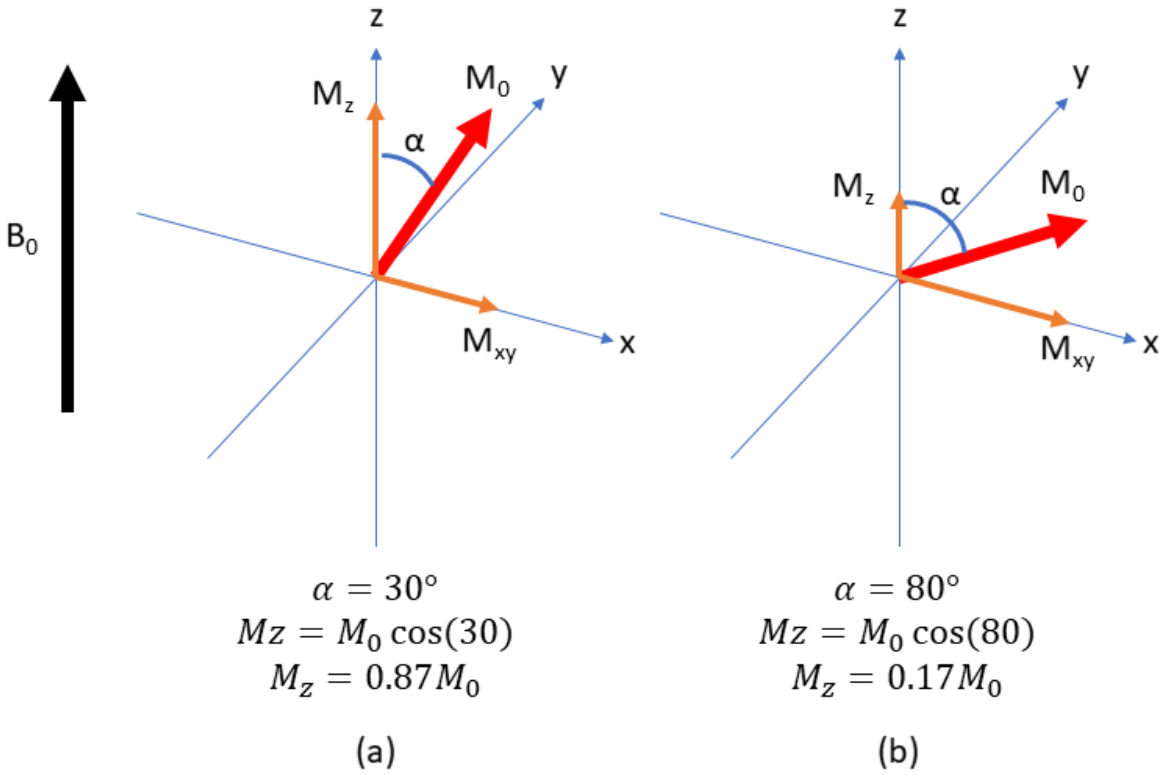



*Figure 2.14: Diagrams showing how the size of $M_z$ changes with flip angle, (a) shows a flip angle of 30° and (b) shows a flip angle of 80°.*

### 2.1.9. Noise and SNR

Thermal noise, which occurs in the receiver coils, is the primary source of noise in MR imaging, and its distribution can be assumed to be Gaussian with zero mean. As well as thermal noise, patient motion, physiological motion and characteristics of the MR system can also introduce noise into the final acquired image (135). In MRI, magnitude images are most commonly presented, which are produced by reconstructing an image from the real and imaginary signals. Magnitude images are preferred as they discard the phase information, which removes any phase artefacts (136). The noise in the real and imaginary signals can be assumed to be Gaussian. However, the process of combining the signals is non-linear, meaning that the final image will have a Rician distribution. However, it has been shown that for SNR values greater than 2, the noise distributions can be assumed to be Gaussian (136).

The simplest method for calculating the SNR in MR images is to calculate the mean signal intensity in a region of interest (ROI) within the brain and divide this by the standard deviation in the signal from an ROI drawn outside of the brain. However, for phased array coils and modern reconstruction techniques this method is not valid and can lead to underestimation or overestimation of SNR (137).

Alternatively, SNR can be calculated following the guidance provided by the National Electrical Manufacturers Association (NEMA) guidelines (138). The first method uses a single acquisition, where an estimate of the noise is calculated by averaging the standard deviation in the signal from four ROIs placed in the four corners of the image. This value is then divided by 0.66, which is to correct for the fact that the background noise follows a Rayleigh distribution instead of a Gaussian distribution. Typically, noise should follow a Gaussian distribution with a zero mean, but because MRI uses magnitude images there are no negative values. This changes the distribution to a Rayleigh distribution with a non-zero mean and a smaller standard deviation. The factor of 0.66 corrects for this reduced standard deviation (128). The second method uses two images which are acquired sequentially with identical acquisition protocol. A subtraction image is produced, and the noise is estimated by calculating the standard deviation in a signal ROI in the subtraction image and dividing by $\sqrt{2}$. The factor of $\sqrt{2}$ is used to to correct for the greater standard deviation in the signal ROI. The factor of 0.66 is not required as no background noise is used in the calculation (128). A third method uses two acquisitions, but in the second acquisition there is no RF excitation, which produces a noise image. The noise can then be estimated from the standard deviation in an ROI in the noise image. If a magnitude image is used, a correction factor of 0.66 must be applied (as is used in the first method). The final method is very similar to the second method as it uses a subtraction image. However, in this case the two images are reconstructed from

the raw k-space data from a single acquisition, which prevents system drift from affecting the noise estimation (138).

Key factors in the SNR of the final acquired image include: the voxel size, $\Delta x \cdot \Delta y \cdot \Delta z$; the relaxation factor $F_{sequence}$; the number of signal averages, NSA; the number of phase encoding steps, $N_{PE}$; the number of frequency encoding steps, $N_{FE}$; and the bandwidth. All of these factors can be related to the SNR through equation 2.16 (128).

$$SNR \propto \frac{\Delta x \cdot \Delta y \cdot \Delta z \cdot F_{sequence} \cdot \sqrt{NSA \cdot N_{PE} \cdot N_{FE}}}{\sqrt{BW}} \tag{2.16}$$

The field strength of the scanner is also important as the SNR of the acquired image is proportional to the field strength. However, increased field strength can have drawbacks such as artefacts appearing more prominently, which is due to an increased artefact-to-noise ratio. An example is the Gibb's ringing artefact, which is a series of parallel lines that can appear at high contrast boundaries (139). Increasing the voxel size increases the SNR. Voxel size is controlled by the field of view (FOV), the slice thickness and the pixel spacing. Increasing the size of any of these factors increases the voxel size. However, increasing the pixel spacing reduces the in-plane resolution and increasing slice thickness increases the chance of partial volume effects (PVEs). Therefore, setting the voxel size is a trade-off between optimal SNR and resolution and reducing PVEs (140). $F_{sequence}$ defines the contribution of the relaxation times to the signal dependent on the type of pulse sequence. The $F_{sequence}$ values for spin echo sequences, $F_{SE}$, and gradient echo sequences, $F_{GE}$, are shown in equations 2.17 and 2.18, respectively (128). In these equations, TR is the repetition time, TE is the echo time, $T_1$ is the $T_1$ relaxation time, $T_2$ is the $T_2$ relaxation time, $T_2^*$ is the $T_2^*$ relaxation time, and α is the flip angle.

$$F_{SE} \propto \left[1 - e^{-\frac{TR}{T_1}}\right] \cdot e^{-\frac{TE}{T_2}} \tag{2.17}$$

$$F_{GE} \propto \frac{\sin(\alpha) \cdot \left(1 - e^{-\frac{TR}{T_1}}\right) \cdot e^{\frac{TE}{T_2^*}}}{1 - \cos(\alpha) \cdot e^{-\frac{TR}{T_1}}} \tag{2.18}$$

Increasing NSA increases the SNR, but also increases the acquisition time (128). Increasing $N_{PE}$ and $N_{FE}$ can increase the SNR. However, increasing the $N_{PE}$ increases the acquisition time. $N_{PE}$ and $N_{FE}$ are also related to voxel volume and if either $N_{PE}$ or $N_{FE}$ are increased without a respective increase in field of view then the voxel size will be decreased (128). The bandwidth refers to the bandwidth of the receiver which records the MR signal. Decreasing bandwidth increases the SNR as less noise is sampled due to the decreased frequency range. However, bandwidth is inversely proportional to the sampling time so decreased bandwidth leads to increased sampling time and therefore longer acquisition time (141).

### 2.1.10 Multi-Slice Imaging

Often when an MR image is required more than one slice is needed. In order to speed up the acquisition time, multiple slices are acquired simultaneously. This can be done using either a 2D multi-slice acquisition or a 3D acquisition. 2D multi-slice imaging exploits the time gap between the end of the signal acquisition and end of the TR. Extra signals can be excited and acquired during this gap allowing for signals from multiple slices to be acquired during a single TR. This is achieved using a series of staggered RF excitation pulses, as shown in Figure 2.15. The number of slices which can be acquired depends on the length of time between the end of the signal acquisition and the next excitation pulse and is therefore dependent on the echo train length and TR. A 3D acquisition uses an extra phase encoding gradient applied along the slice select direction. The slice select gradient and RF excitation pulse are used to excite a 3D volume, which is then partitioned into slices using the extra

phase encoding gradient. Figure 2.16 shows the difference in slice select gradient between a 2D and 3D gradient echo pulse sequence. A 3D acquisition can only be carried out using gradient echo or fast spin echo imaging, as standard spin echo imaging is too slow. However, it offers thinner slices than 2D multi-slice and better SNR for comparable slice thickness, and images produced from a 3D acquisition can be reconstructed into any plane provided the voxels are isotropic. As a general rule, a 2D acquisition is used if the slice thickness is 3mm or greater, whilst a 3D acquisition is used for thinner slices (128).

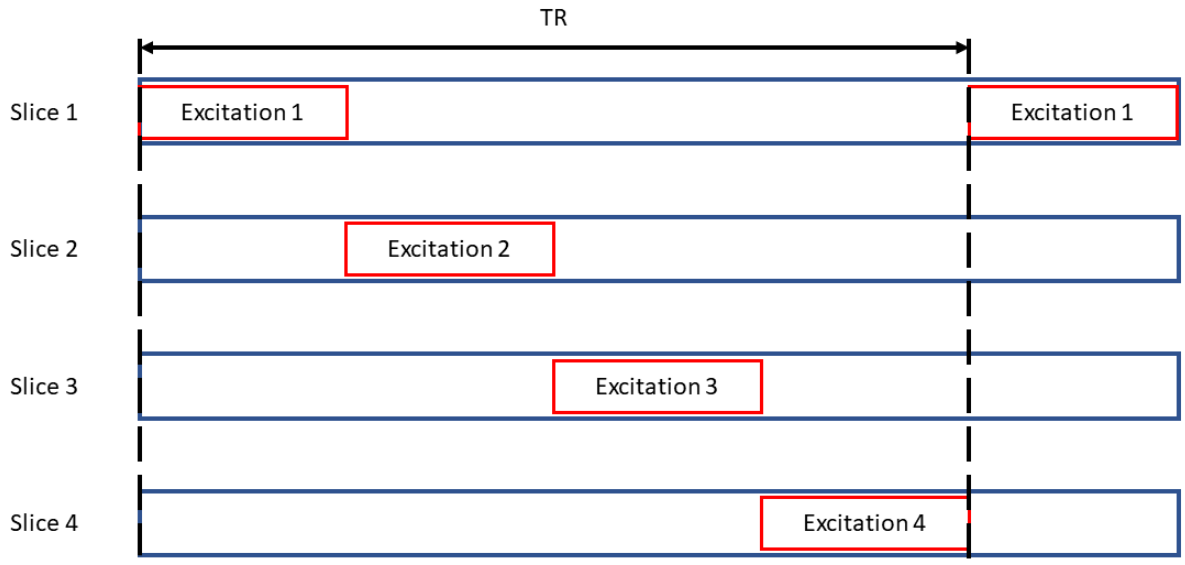



*Figure 2.15: An example of how staggered RF excitation pulses are applied in a 2D multi-slice acquisition.*

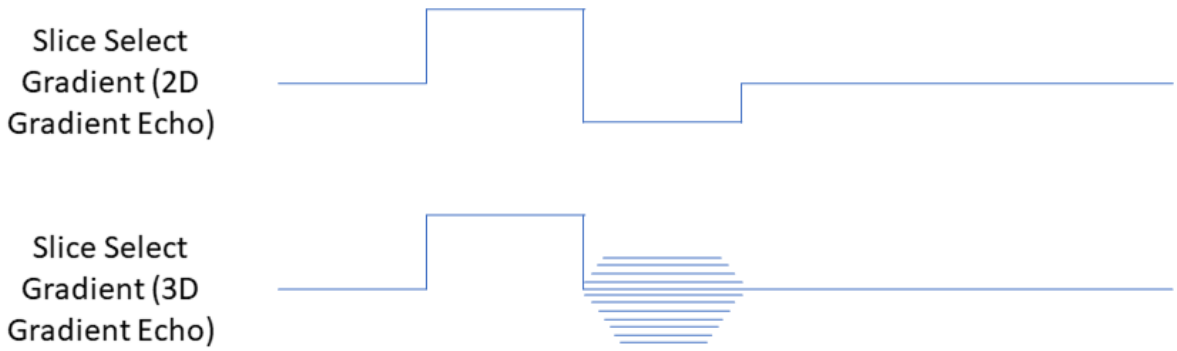



*Figure 2.16: The difference in the slice select gradients between a 2D and 3D gradient echo acquisition. The 3D version includes phase encoding gradients to allow multiple slices to be acquired simultaneously.*

## 2.2. DSC-MRI

A brief overview to DSC-MRI, its applications and issues with the method was given in the introduction. This section will explain the physics behind estimating perfusion with a contrast agent, and how DSC-MRI data is acquired and how it is analysed to produce biomarkers.

### 2.2.1. Estimating Perfusion with a Contrast Agent

DSC-MRI relies on the use of a paramagnetic contrast agent, which is injected into the blood. Paramagnetic contrast agents consist of molecules that have their own magnetic moment, which is either caused by unpaired nucleons in the nucleus, or unpaired electrons orbiting the nucleus (142). A contrast agent must meet several criteria to be used in DSC-MRI: there should be no chemical interactions between the contrast agent and the blood plasma; the contrast agent must not disturb the blood flow; it must be detectable by the imaging method; and finally it needs to be non-diffusible (143).

Gadolinium-based contrast agents are frequently used in DSC-MRI, as gadolinium has a large magnetic moment due to having 7 unpaired electrons. However, gadolinium is toxic so must be chelated with a ligand before it can be safely injected. Commonly, it is chelated with diethylene-triaminepenta-acetic acid (DTPA) to form Gd-DPTA (144). Gd-DPTA has $T_1$ and $T_2$ relaxivities of 4.3 $Lmmol^{-1}s^{-1}$ and 4.4 $Lmmol^{-1}s^{-1}$, respectively, when measured in blood at 1.5T (30). It causes signal changes in MR imaging by shortening $T_1$ and $T_2$ values, which accelerates the longitudinal and transverse relaxation. The magnetic susceptibility of the contrast agent also increases the difference in magnetisation between the blood and surrounding tissues. This causes local inhomogeneities in the magnetic field, leading to local variations in $T_2^*$. Reduced local $T_2$ and $T_2^*$ cause local signal decreases in $T_2$ and $T_2^*$ weighted imaging, respectively (143).

The changes in signal intensity caused by the contrast agent can be plotted as a signal-time course, and each pixel within a DSC-MRI image has a corresponding signal-time course. An assumption of DSC-MRI is that the change in concentration of gadolinium contrast agent, represented by a concentration time course (CTC), is linearly proportional to the change in $T_2^*$ relaxivity, $\Delta R_2^*$, caused by the contrast agent, which is calculated from the signal-time course. These quantities are all related through Equation 2.19, where $C_t(t)$ is the CTC, $\Delta R_2^*(t)$ is the change in $\Delta R_2^*$ with time, $\Delta T_2^*$ is the change in $T_2^*$ relaxation time, TE is the echo time, S(t) is the signal at time, t, obtained from the signal-time course, and $S_0$ is the mean of the baseline of S(t) (145). Figure 2.17 shows an example of a signal-time course and $\Delta R_2^*$ curve.

$$C_t(t) \propto \Delta R_2^*(t) = \frac{1}{\Delta T_2^*} = -\frac{1}{TE}\ln\left(\frac{S(t)}{S_0}\right) \qquad (2.19)$$

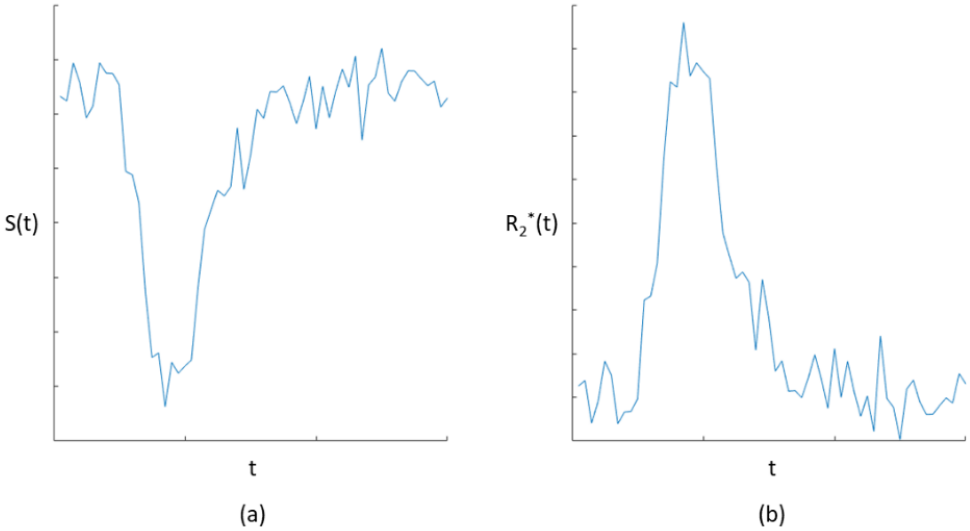



*Figure 2.17: An example signal-time course (a) and its corresponding relaxation rate curve (b).*

### 2.2.2. Acquisition

The acquisition protocol for DSC-MRI can be divided into two parts: the imaging protocol and the injection protocol for the contrast agent. In order to monitor the changes in concentration of the contrast agent the DSC-MRI images are acquired rapidly. Currently, it is recommended that images are acquired with a temporal resolution of at least 1.5s (34). For

this reason, the recommended protocol for acquiring DSC-MRI data is the gradient echo – echo planar imaging (GE-EPI) sequence (34). However, it is also possible to acquire DSC-MRI using the sensitivity encoded (SENSE) principles of echo-shifting with a train of observations (sPRESTO) (146). Both protocols need to be $T_2^*$ weighted, as changes in $T_2^*$ caused by the contrast agent are being measured (147). It is also possible to acquire DSC by measuring changes in $T_2$ using a spin echo pulse sequence. However, GE-EPI has advantages over spin echo such as greater SNR and sensitivity (34).

#### 2.2.2.1. Imaging Protocol

Both the GE-EPI and sPRESTO sequences have RF excitation pulses with flip angles, α, which are less than 90°. The GE-EPI sequence, shown in Figure 2.18, differs from standard gradient echo imaging in that it uses a train of frequency encoding gradients to create a train of gradient echoes following one RF pulse. As each frequency encoding gradient in the train has an individual phase encoding gradient, multiple lines of k-space can be acquired from a single RF pulse, which speeds up acquisition (131). For GE-EPI it is common to use a spatial-spectral (SPSP) RF excitation pulse, which is used to reduce lipid induced chemical shift artefacts. This consists of an envelope of RF pulses applied at the same time as an alternating slice select gradient. The frequencies within the RF envelope determine the spectral content of the signal, whilst the combination of the slice select gradient and RF frequencies determine the spatial selectivity. An oscillating slice select gradient is required to prevent the production of spatially offset replicates of the slice profiles for each of the chemical species (131).

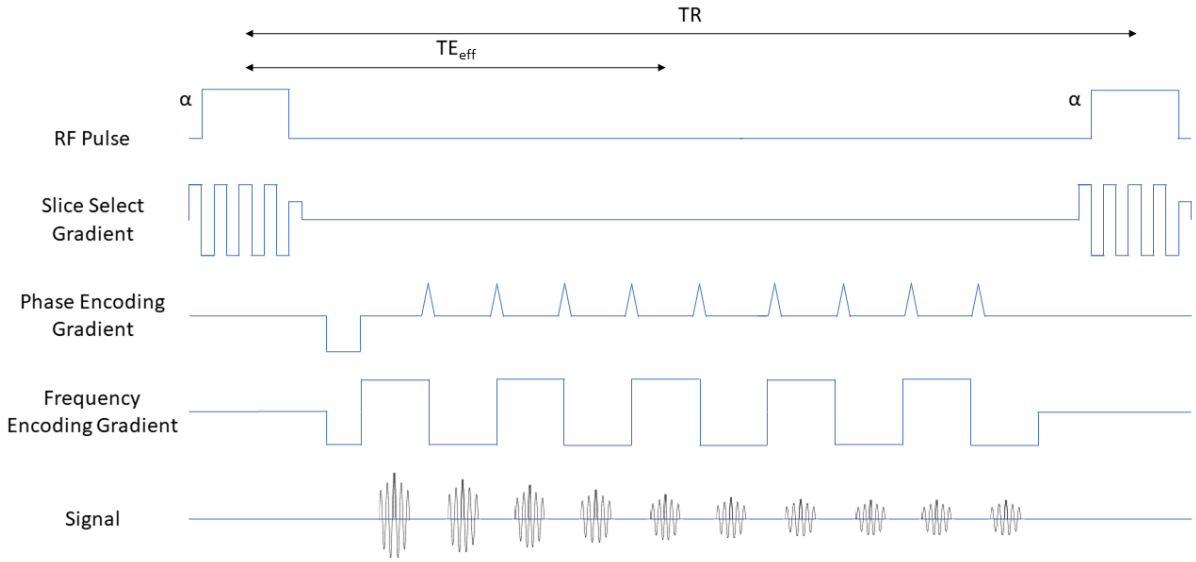



*Figure 2.18: GE-EPI pulse sequence diagram.*

The sPRESTO sequence, shown in Figure 2.19, also uses an echo train, but in combination with an echo shifted pulse sequence. In an echo shifted pulse sequence a modified slice select gradient is used to ensure that no signal is recorded in the first TR following the RF pulse, which delays the gradient echo by one TR. Following the first TR, the slice select gradient consists of a positive lobe of area, A, followed by two lobes of equal and opposite area, 2A (shown in Figure 2.19). The first positive lobe rephases the gradient echo from the first TR and the pair of lobes that follow it spoil the transverse magnetisation from the second TR (131).

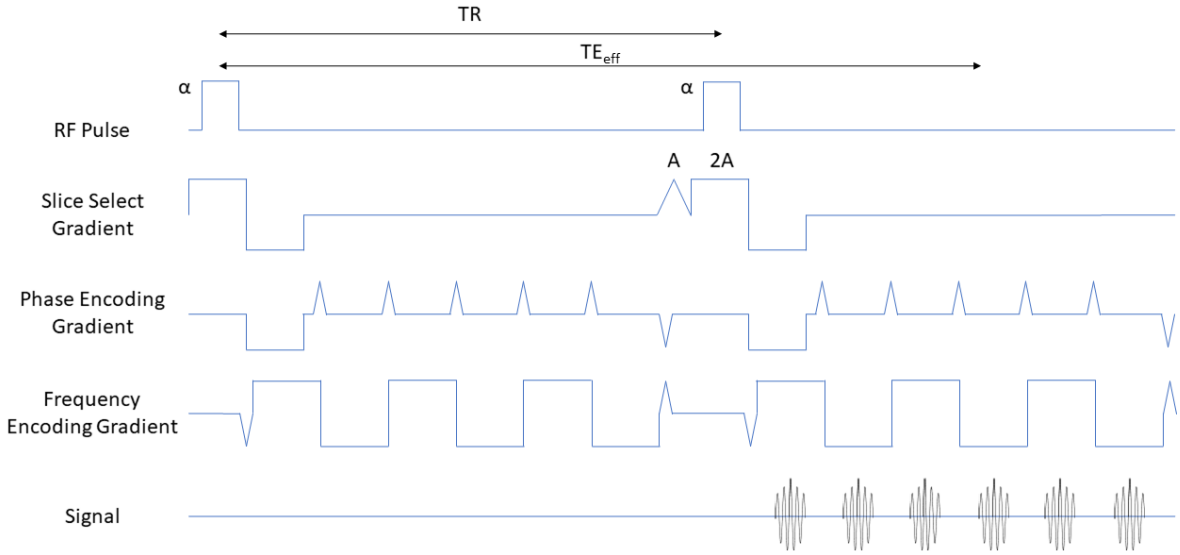



*Figure 2.19: sPRESTO pulse sequence diagram. No signal is recorded during the first TR and is rephased using the slice select gradient, which consists of a positive lobe of area A, followed by a positive and negative lobe, each with area 2A.*

A factor that needs to be considered with DSC-MRI acquisition is the length of the TR relative to the $T_1$ and $T_2$ values of the tissues imaged. When the TR is significantly longer than the $T_1$ and $T_2$ values, the transverse magnetisation completely decays between successive RF pulses. However, if the TR is shorter than the $T_1$ and $T_2$ values then there is residual transverse magnetisation at the end of the TR, which experiences the next RF pulse. Therefore, it takes a few repetitions of the pulse sequence before a steady state is reached, where the net transverse magnetisation is equal at the same point in two successive pulse sequences (148). This is particularly a problem for sPRESTO as the TR is very short and is in the order of ms. For sPRESTO scans it is common to acquire a few dummy scans, where the pulse sequence is played but no signal is acquired, to allow the steady state to be reached (149, 150).

The total acquisition time for a MR image is defined by the length of TR, the number of NSAs and the number of repetitions of TR needed to acquire enough lines of k-space, as shown in equation 2.15. Reducing the acquisition time increases the temporal resolution. The temporal resolution is important as it needs to be short enough to be able to capture the

change in signal intensity induced by the contrast agent. Currently a temporal resolution of 1.5s or less is recommended (34). $T_2^*$ weighted GE-EPI and sPRESTO imaging techniques are used to acquire DSC-MRI as both techniques allow multiple lines of k-space to be acquired with a single excitation pulse, offering much faster acquisition, and therefore greater temporal resolution, than standard gradient echo imaging (146).

Parallel imaging can be used to further reduce acquisition time, by undersampling k-space in the phase encoding direction and reconstructing an image from the undersampled data. The reduction in acquisition time is determined by the acceleration factor, R, which is a ratio of the number of lines of k-space acquired in the fully sampled image, to the number of lines of k-space acquired in the undersampled image. Parallel imaging uses a phased coil array, consisting of several receiver coils. Each coil is more sensitive to the region of tissue it is closest to, providing additional spatial information which can be used in the reconstruction (151).

Undersampling k-space can result in aliasing artefacts, which can be removed by parallel imaging reconstruction algorithms. There are many reconstruction algorithms available, and they can be classified as either sensitivity encoding (SENSE) algorithms or generalised autocalibrating partially parallel acquisition (GRAPPA) algorithms. SENSE algorithms combine the aliased images with a map of the coil sensitivities to produce a final, unaliased image. GRAPPA algorithms are applied to the raw k-space data and regenerate the missing phase-encoding lines (151).

When optimising the acquisition parameters for DSC-MRI, there are a series of trade-offs between improving the data quality, maintaining sufficient temporal resolution, and reducing any $T_1$ effects in the acquired data. $T_1$ effects can contaminate the $T_2$ / $T_2^*$ weighted DSC-

MRI signal due to $T_1$ shortening caused by the gadolinium contrast agent leaking from the intravascular space to the extravascular extracellular space (EES) in tumours with a disrupted blood brain barrier (BBB). These effects can be reduced by reducing the $T_1$ weighting of the sequence - reducing the flip angle and increasing the TE. However, reducing the flip angle reduces SNR, and increasing the TR reduces the temporal resolution, which increases the acquisition time (34).

#### 2.2.2.2. Injection Protocol

A gadolinium based contrast agent is injected during acquisition using a power injector. The injection rate varies between patients and depends on factors such as the age of the patient and the cannula size (34). This is particularly important in paediatrics where custom bolus administrations are needed (152). Contrast agent may be injected either as a single-dose dynamic bolus, or as a double-bolus consisting of the dynamic bolus and a preload. The American Society of Functional Neuroradiology (ASFNR) recommendations for DSC-MRI from 2015 suggest a preload of a quarter to a single dose, 5-10 minutes prior to the injection of the main dynamic bolus should be used (34). However, for paediatrics European society for paediatric oncology (SIOPE) recommends that the total bolus is equally split between the pre-bolus and the main bolus, in order to limit the amount of contrast agent injected (153). The selection of preload dose and time between preload and main dynamic bolus is important as it has been shown to have an effect on the accuracy of the CBV values (154-158).

In patients with a disrupted BBB, leakage of contrast agent from the intravascular space causes $T_1$ and $T_2$ or $T_2^*$ weighted effects, which affects the accuracy of the acquired signal and therefore the accuracy of any biomarkers estimated from the data (34). The purpose of the preload is to leak from the intravascular space to the extravascular extracellular space (EES),

thereby shortening the $T_1$ of the EES prior to the DSC acquisition, minimising the $T_1$ weighted effects that occur during the acquisition of the DSC-MRI signal (59).

#### 2.2.2.3. Recommended Acquisition Protocol

In 2015 the ASFNR published a set of recommendations for acquiring DSC-MRI data (34). A more updated set of recommendations were published in a consensus paper in 2020 (159). Table 2.1 summarises the two set of recommendations and the differences between them. The main difference between these two protocols is the flip angle and the use of a preload. As mentioned above the purpose of the preload is to reduce $T_1$-weighted effects cause by contrast agent leakage. However, recent work has shown that using a lower flip angle and no preload produces similar results, provided that leakage correction is applied (160, 161).

| Acquisition Parameter | Paper | |
|---|---|---|
| | ASFNR 2015 Recommendations | 2020 Consensus Paper |
| Sequence | GE-EPI | GE-EPI |
| TR (ms) | 1000-1500 | 1000-1500 |
| TE (ms) | 25-35 | 25-35 |
| Flip Angle | 60˚ - 70˚ | 30˚ |
| Temporal Coverage | 120 time points, with injection of contrast agent 30-50 time points after imaging begins | At least 120 time points, with injection of contrast agent 30-60 time points after imaging begins |
| Slice Thickness (mm) | 3-5 | 3-5 |
| Field of View (cm) | 20 | 24 |
| Preload | 1/4 - 1 single dose 5-10 mins prior to DSC acquisition | No preload |
| Injection Rate (mL/s) | 3-5 | 3-5 |
| Total Acquisition Time | Approximately 2 minutes | 2-3 minutes |

*Table 2.1: Summary of the protocols presented in both the 2015 and 2020 consensus papers.*

### 2.2.3. Analysis of DSC-MRI

Before analysis of DSC-MRI data can be carried out, it is necessary to assess the quality of the raw data, which is the signal-time courses produced by the acquisition. Currently the

ASFNR recommendation is that this is carried out by visual inspection of the data, which involves assessing whether a clear first pass can be seen and assessing the level of noise within the baseline and the rest of the signal (34). This is an important step as the quality of the raw data affects the accuracy of any metrics calculated from the signal-time courses and any biomarkers calculated from the concentration time curves. The quality of the raw data can be affected by a series of artefacts, which are discussed in Section 2.2.4.

#### 2.2.3.1 Metrics from DSC-MRI Signal-Time Courses

Metrics which can be calculated from DSC-MRI signal-time courses include signal drop to noise ratio (SDNR), root mean square error (RMSE), full width half maximum (FWHM) and Percentage Signal Recovery (PSR). SDNR can be calculated using equation 2.20, with the signal drop defined as the difference between the mean baseline and mean of the first pass minima and the two adjacent dynamics. RMSE can be calculated by fitting a version of the simplified gamma variate function (162), shown in equation 2.21, to the first pass of the signal-time course. In equation 2.21 y(t) is the fit, t is the time, c is the average baseline, and α, β and K are coefficients that affect the shape of the fit. The FWHM can be determined from the width of the bolus of the signal-time course (in seconds) at half of the signal drop. The PSR can be calculated from equation 2.22 (34), with $T_2^*$ recovery defined as the difference between the mean post-bolus signal and the mean of the first pass minima and the two adjacent dynamics.

$$SDNR = \frac{Signal\ Drop}{Standard\ Deviation\ in\ Baseline} \tag{2.20}$$

$$y(t) = c - Kt^{\alpha}e^{-\frac{t}{\beta}} \tag{2.21}$$

$$PSR = \frac{T_2^*\ Recovery}{Signal\ Drop} \times 100 \tag{2.22}$$

#### 2.2.3.2 Estimating Biomarkers

Several biomarker values can be estimated from DSC-MRI including: CBV, cerebral blood flow (CBF), vascular mean transit time (MTT), bolus arrival time (AT) and time to peak (TTP). AT is defined as the time between the bolus injection and the beginning of the signal drop, whilst TTP is the time between the bolus injection and the maximum signal drop (163). CBV, CBF and MTT values are calculated for each voxel by converting the signal-time course into a CTC, using equation 2.19 (156).

A relative CBV (rCBV) value can be calculated by fitting a gamma variate to the first pass of the CTC and calculating the area under the curve (147). rCBV is a commonly reported measure in DSC-MRI research and is often normalised using an average rCBV value from an ROI defined in normally appearing white matter (WM) or grey matter (GM) (34).

In order to obtain fully quantified CBV values and to obtain CBF and MTT values it is necessary to obtain an AIF. An AIF is a CTC obtained from a voxel located within an artery, it therefore represents the concentration of contrast agent within the blood vessels. The artery selected must be supplying the tissue which the CBV is being calculated for. A quantified CBV value can be obtained from dividing the area under the CTC by the area under the AIF, as shown in equation 2.23, where $C_t(t)$ is the CTC, $C_a(t)$ is the AIF, $H_f$ is a factor accounting for the differences in haematocrit between large and small vessels, and ρ is the brain tissue density (147).

$$CBV = \frac{H_f}{\rho} \frac{\int_0^t C_t(t)dt}{\int_0^t C_a(t)dt} \tag{2.23}$$

The CTC for each voxel can be assumed to be a convolution of the AIF and a tissue residue function, R(t), all multiplied by the CBF, as shown in equation 2.24 (147). The residue

function describes the amount of contrast agent remaining in the tissue following the injection of the bolus of contrast agent (147).

$$C_t(t) = CBF \cdot C_a(t) \otimes R(t) \tag{2.24}$$

Deconvolution of the AIF and CTC allows a quantified value of CBF to be calculated. There are multiple methods for performing the deconvolution, but the most frequently used is singular value decomposition (SVD). SVD is a technique for factorising matrices and is explained in section 2.3.1. The vascular MTT for the voxel can then be calculated from the central volume theorem, which relates the MTT, CBF and CBV and is shown in equation 2.25 (147).

$$MTT = \frac{CBV}{CBF} \tag{2.25}$$

In most clinical applications AIF selection is often manual and relies on a radiologist drawing a ROI around one of the major arteries (often either the major carotid artery or the internal carotid artery). However, manual AIF determination can be time consuming and subjective between radiologists, and also relies on the existence of an appropriate vessel within the FOV (164). Therefore, most centres do not carry out deconvolution and just report rCBV values instead. There are several automated AIF selection methods available (164-167), but their use is not widespread.

### 2.2.4. DSC-MRI Artefacts

Artefacts such as noise, motion, susceptibility artefacts and contrast agent leakage, make it challenging to obtain accurate biomarkers from the acquired DSC-MRI data (34, 168). This section covers each of these artefacts and the methods used to reduce or correct them.

#### 2.2.4.1 Noise

Noise is a common problem across all MR modalities. In DSC-MRI, noise can obscure the first pass of the signal-time course, making it difficult to obtain accurate estimates of biomarker values. The acquired DSC-MRI signal is affected by the field strength and the acquisition parameters used. If the signal is increased but the noise remains constant then the SNR of the acquired data is improved (34).

Alternatively, noise can also be reduced in post-processing by denoising the acquired data. Examples of denoising methods include: matricization, Tucker decomposition and wavelet denoising. The theory behind these methods is covered in section 2.3.

#### 2.2.4.2 Motion

Patient motion can introduce noise and discontinuities into the acquired DSC-MRI signal-time course. These discontinuities result in a spatiotemporal partial volume effect (PVE), which effects the accuracy of biomarker estimates obtained from the acquired data. This type of PVE also makes AIF determination challenging, as the arteries AIFs are selected from are often around a voxel in diameter. This makes it more difficult to obtain quantified biomarker values (169).

Motion can be reduced by acquiring the DSC-MRI scan whilst the patient is under anaesthetic, which is common practice in paediatric studies where the patients are 5-6 years old or younger (152, 170, 171). It is also possible to correct for motion in post-processing by registration methods, with one image from one time point used as a reference to register the rest of the time points to (169).

#### 2.2.4.3 Susceptibility Artefacts

Susceptibility artefacts are distortions to an image caused by inhomogeneities in the $B_0$ field, which are in turn caused by variations in the magnetic susceptibility. These distortions

generally occur at boundaries between air and tissue (for example the sinuses) and around metal implants (172).

Susceptibility artefacts can be reduced by optimising the imaging protocol to increase the pixel bandwidth in the phase encoding direction (172). It is also possible to correct for susceptibility artefacts using post-processing correction, and there are several different methods available. For example, in one method two GE images with differing TEs are acquired, and the phase differences between the images are used to estimate the inhomogeneity (173). Another method is to acquire two gradient echo (GE) images, where one has a reversed phase encoding gradient. The two images then have opposite distortions, which can be used to estimate an undistorted image (172, 174).

#### 2.2.4.4 Contrast Agent Leakage Correction

Contrast agent leakage from the intravascular space to the extracellular extravascular space occurs in patients with a disrupted BBB. This results in $T_1$ and $T_2^*$ weighted effects that can affect the accuracy of biomarker values from DSC-MRI data. In regions where $T_1$ weighted effects dominate, contrast extravasation causes a decrease in $T_2^*$ weighted signal loss, which results in an increased post-bolus signal in the signal-time course, leading to underestimated CBV values. In other regions differences in magnetic susceptibility between the contrast agent in the EES and the surrounding tissue can lead to $T_2^*$ weighted effects dominating, which leads to a reduced post-bolus signal in the signal-time course and overestimated CBV values (59). Figure 2.20 shows an example of the $T_1$ and $T_2^*$ weighed effects on a signal-time course.

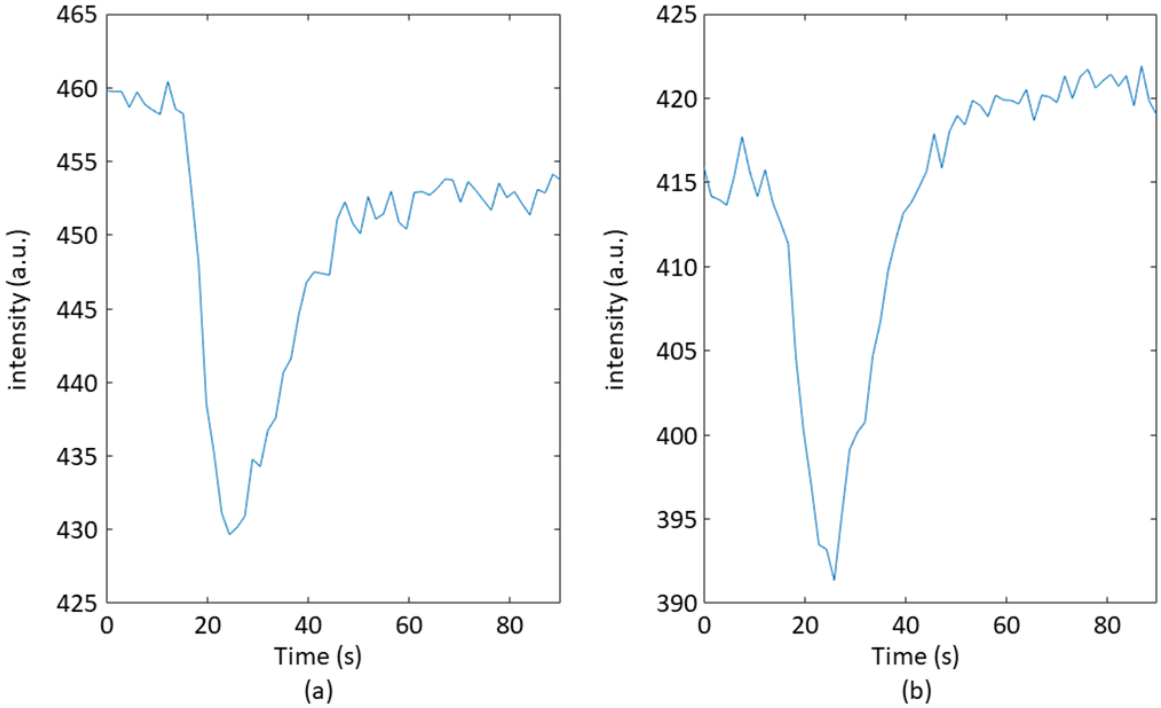



*Figure 2.20: Examples of the effects of contrast agent leakage on signal-time courses. (a) shows an example of $T_2^*$ weighted effects, whilst (b) shows an example of $T_1$ weighted effects.*

There are several different leakage correction methods available (156, 175-178). However, the most widely used method is the Boxerman method (156). In this method the leakage correction is applied to the uncorrected CTCs, $\Delta\bar{R}_2{}^*$, associated with each signal-time course. The uncorrected CTC can be linked to the corrected CTC, $\Delta R_2{}^*$, through equation 2.26, where $\overline{\Delta R_2^*}$ is the average CTC for voxels with no contrast agent leakage, and $K_1$ and $K_2$ are correction factors. $K_1$ multiplied by $\overline{\Delta R_2^*}$ represents the corrected $\Delta R_2{}^*$, and $K_2$ represents the effects of contrast agent leakage (156). $\overline{\Delta R_2^*}$ is estimated by finding all CTCs where the mean of the last 10 time points are within 1 standard deviation of the average baseline (156).

$$\Delta\bar{R}_2{}^*(t) \approx K_1\overline{\Delta R_2^*}(t) - K_2\int_0^t \overline{\Delta R_2^*}(t')dt' \qquad (2.26)$$

This equation makes several assumptions. Firstly, it assumes that $R_2{}^*$ is approximately proportional to the CTC (equation 2.19). Secondly, it assumes that the effect of contrast agent

leakage is a small and occurs over a time period longer than the first pass, but shorter than the clearance time of the contrast agent. Thirdly, it assumes that the contrast agent only diffuses from the intravascular space to the extravascular extracellular space. Finally, it assumes that the average of CTC's that do not demonstrate leakage are proportional to the average contrast agent in the capillaries, and that they all return to the baseline (156).

As shown in equation 2.27, $K_2$ is dependent on a number of factors including: TR; TE; $T_1$; the longitudinal relaxivity of the contrast agent, $r_1$; a constant, k; and the permeability and surface area product, PS. The constant k depends on physiological factors such as vessel size and the average CBV within the brain. PS is an important parameter as it is the only leakage dependent parameter in the equation for $K_2$, and therefore explains why $K_2$ defines the amount of leakage (156). A Similar measure to PS in DCE-MRI is $K^{trans}$ which measures a mixture of PS and flow (179).

$$K_2 = \frac{TR}{TE}\frac{e^{-\frac{TR}{T_1}}}{1 - e^{-\frac{TR}{T_1}}} r_1 \cdot k \cdot PS \tag{2.27}$$

Since the value of $\overline{\Delta R_2^*}(t)$ is constant for each voxel, the values for $K_1$ and $K_2$ can be estimated using a linear least squares fit using equation 2.26. The corrected $\Delta {R_2}^*$ can then be calculated from rearranging equation 2.26, and a corrected value or CBV can be calculated using equation 2.28. In this equation rCBV is the uncorrected relative CBV value and $rCBV_{corr}$ is the corrected rCBV value (156).

$$rCBV_{corr} = rCBV + K_2 \int_0^{\mathrm{T}} dt'' \int_0^{t''} \overline{\Delta R_2^*}(t')dt' \tag{2.28}$$

This method has also been further extended by Leu et al. to allow for bidirectional leakage corrections (180).

## 2.3. Computational Techniques

Developing a pipeline for processing DSC-MRI data requires multiple computational techniques to process the acquired data. This section provides some background on the computational techniques used. It covers the singular value decomposition (SVD) and wavelet based methods which are applied as denoising methods. This is followed by machine learning techniques including: principal component analysis (PCA) and clustering techniques, used for segmentation; and classification, used for determining data quality. Finally, section 2.3.5 gives an overview of the types of data used in this thesis and examples of how each computational technique is suited to the data.

Machine learning techniques can be classified as either supervised or unsupervised learning. In supervised learning techniques the training data consists of a set of input data and a set of corresponding correct outputs. In unsupervised learning the training data consists of only input data with no outputs (181). Both supervised and unsupervised machine learning techniques are used in this thesis. The supervised machine learning techniques have been used for classification, whilst the other unsupervised techniques have been used for denoising, dimensionality reduction, and segmentation.

### 2.3.1. Singular Value Decomposition

The singular value decomposition (SVD) is a technique for factorising a matrix into a series of constituent matrices. SVD is used to estimate the rank of matrices and to calculate a lower-rank approximation of a matrix (182). It also provides the basis for other techniques including principal component analysis (PCA) and the Tucker decomposition (183). It states that a matrix, A, can be factorised into three matrices, according to equation 2.29. A is a matrix of dimensions $m \times n$, whilst U and V are orthogonal matrices of dimensions $m \times m$ and $n \times n$ respectively and S is a diagonal matrix with the same dimensions as A (184).

$$A = USV^T \tag{2.29}$$

A lower rank approximation of the original matrix can be estimated by rewriting the matrix, A, as the sum of a series of rank 1 matrices, according to equation 2.30, where $s_{ii}$, $u_i$, and $v_i$ are all elements of S, U and V from equation 2.29 respectively, and r is the number of elements in the diagonal of S (185).

$$A = \sum_{i=1}^{r} s_{ii} u_i v_i^T \tag{2.30}$$

The $s_{ii}$ values are the diagonal values of S and are referred to as the singular values. The p largest singular values (where $p < r$) can be selected and used to produce a reconstruction of A, referred to as, and shown in equation 2.31 (185).

$$A_S = \sum_{i=1}^{p} s_{ii} u_i v_i^T \tag{2.31}$$

Assuming the largest singular values are associated with the signal and the smallest singular values are associated with the noise, this has the effect of removing the noise from the reconstructed dataset (185).

Rank reduction, via SVD or SVD-based methods, has been widely applied as a denoising method in the literature. It has been applied to imaging: Guo et al. and He et al. have both demonstrated that SVD denoising can be used to suppress the noise in an image and gives improved performance when compared to spatial domain filtering methods (186, 187); James et al. has applied three SVD-based to MR images, with all three methods improved the SNR (188); Lyra-Leite et al. have demonstrated that it can be applied to MRI images to reduce noise and make the reconstruction process less computationally expensive (189); and Leal et

al. have applied it to MR images and demonstrated improved SNR and reduced artefacts compared to other spatial filtering methods (190).

It has also been applied to a range of signals: Cheng et al. have shown it can be applied to laser imaging, detection, and ranging (LIDAR) data to give improved SNR (191). Fan et al. have applied it to denoise Raman spectroscopy data, where it was shown to provide a more efficient alternative to standard methods of denoising individual spectra (192); Li et al. have shown that it can be applied to ECG data (193); Wang et al. have applied it to unmanned air vehicle flight data, in combination with standard analysis methods to give improved noise suppression (194); and Jha et al. have applied it to give good noise suppression in electronic nose data (195).

#### 2.3.1.1. Principal Component Analysis

Principal component analysis (PCA) is a form of unsupervised machine learning, which can be thought of as a method of reorganising a dataset into a new set of variables. This new set of variables are called principal components (PCs) and are used to explain the variance in the dataset. The first PC is oriented in a direction to explain the maximum variance. All following PCs are then orthogonal to the first PC and all other PCs. All the data within the dataset will have a new set of values for the PCs. These new values are referred to as scores and each point within the dataset will have a score for each PC (196).

PCA uses SVD (equation 2.29) to calculate the principal components. In this case the columns of the matrix, A, are variables and the rows are the corresponding observations. Often the columns of A are centred prior to PCA, so that all columns have a zero mean. Following SVD the orthogonal matrix, V, is referred to as the principal component coefficient matrix and indicates the correlations between the principal components and the variables. The

principal component score matrix, F, can then be calculated using equation 2.32, where U and S are matrices produced by SVD (equation 2.29) (197).

$$F = US \tag{2.32}$$

The aim of PCA is to reduce the dimensionality of the dataset. The square of the singular values is equivalent to the eigenvalues of the covariance matrix. The magnitude of each eigenvalue is proportional to the amount of variance explained by its corresponding principal component. Therefore, the singular values contain information on the amount of variance explained by each principal component. Ordering the singular values and their corresponding principal components by descending order ensures the principal components are ordered by the amount of variance they explain (197). This means that by rearranging the data into a set of PCs, it is possible to represent the dataset with fewer variables with a minimal loss of information (198).

Representing a dataset with fewer variables means that PCA analysis has applications in exposing relationships between variables that might not be apparent in a highly dimensional dataset, and in denoising data (198). One of the challenges of applying PCA as an analysis technique is selecting the number of principal components to use. A commonly used technique is the scree plot, where the eigenvalues are plotted against the components (199). An example scree plot is shown below in Figure 2.21. The number of components to include is then selected by finding the ‘elbow’ of the curve where it starts to level off and any additional components have low variance (200). There are however, many different methods available for selecting the number of principal components, some which use a variation on the Scree plot (200) and others which use different methods (201-203).

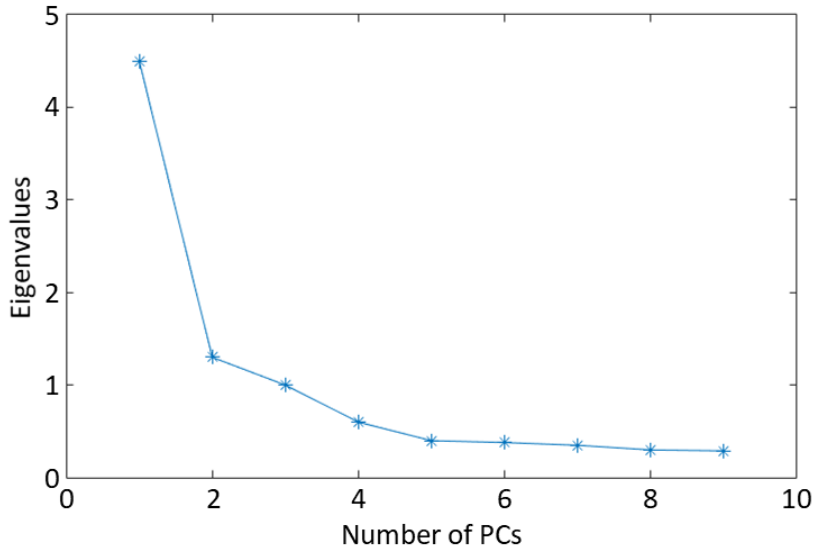



*Figure 2.21: An example of a scree plot used to select the number of principal components used in PCA analysis*

PCA has previously been applied as a dimensionality reduction method in multiple applications. Perfilieva and Hurtik have applied it as a technique to aid pattern recognition in large databases (204), whilst Zhao et al. have applied it as a dimensionality reduction technique to improve the performance of a deep learning network used for image classification (205). In MRI it has been applied by Kaya et al. in combination with clustering methods to segment brain tumours from $T_1$ weighted MR images (206). It has also been applied as a method to aid the analysis of functional MRI (fMRI) (207).

PCA also has multiple applications as a denoising method and can be applied to images and time-varying signals. James et al. applied it to denoise grayscale and colour images and showed that it gave improved SNR (188), whilst Malladi et al. compared a PCA method to standard image denoising methods and showed that PCA gave improved performance (208). In terms of time-varying signals, Jade et al. have demonstrated that PCA can be used to recover the original signal (209), whilst Kang et al. have applied it to decompose multi-channel electrocardiogram (ECG) signals during the denoising process (210). In MRI, Manjon et al. and Kanwal et al. have both applied it to denoising of MRI images (211, 212),

whilst Gurney-Champion et al have applied it to diffusion weighted MRI (201), Zhu et al have applied it to ASL (213), and Balvay et al have applied it to DCE-MRI (214).

#### 2.3.1.2. Matricization

SVD and PCA methods are limited by the fact that they can only be applied to a 2-dimensional matrix. However, many datasets have more than 2 dimensions. Often in imaging this will be two spatial dimensions, and one signal or data dimension. Matricization is a technique which allows a higher order tensor to be unfolded into a matrix, allowing for techniques such as PCA to be applied to the data. The tensor can be unfolded along any of its dimensions, for example a 3 x 4 x 2 tensor can be unfolded into 3 matrices: two 3 x 8 matrices and a 2 x 12 matrix, as demonstrated in figure 2.22 (183). A detailed explanation of the formulas used to map the tensor indices to the matrix indices can be found in work by Kolda (215).

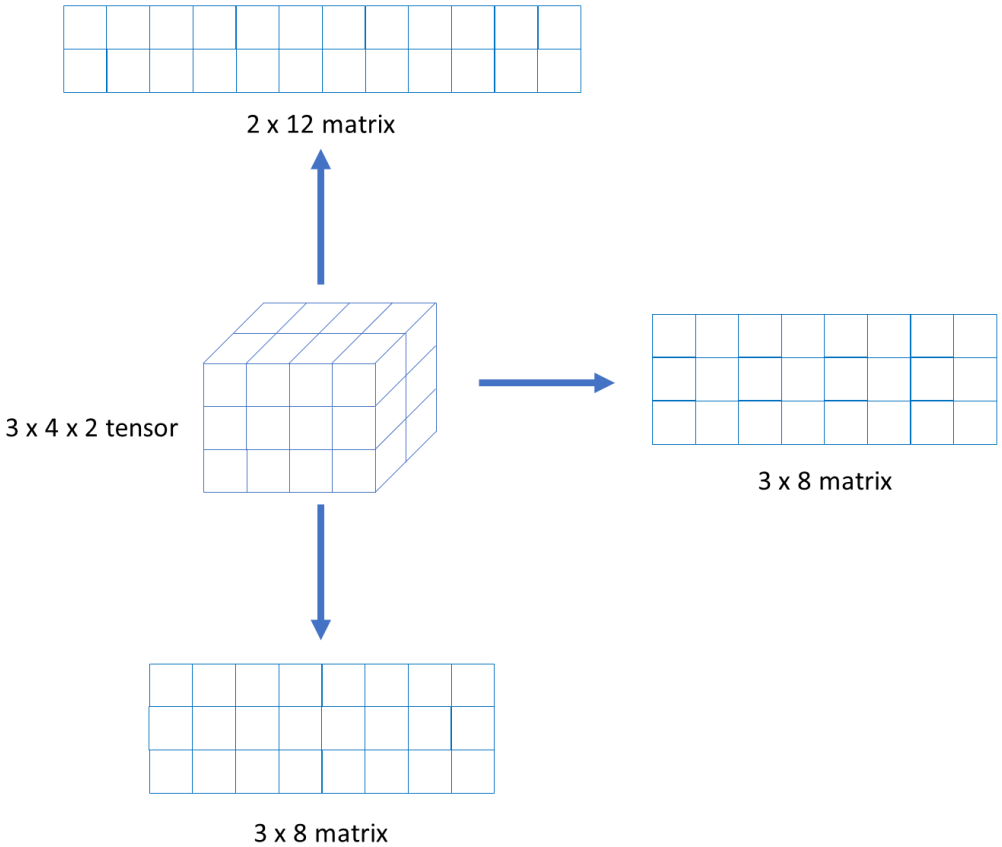



*Figure 2.22: A demonstration of how a 3 x 4 x 2 tensor can be unfolded into matrices along its three dimensions.*

Following matricization SVD-based denoising can be applied to the data and then the matrix can then be refolded into a tensor (216). Matricization has been applied in the literature in combination with SVD-based denoising method. For example, Brender et al. have used it to denoise MR spectroscopy data (216), and Feschet has used matricization in the denoising of images (217). However, the main application of matricization is in tensor decomposition (183).

#### 2.3.1.3. Tucker Decomposition

The Tucker decomposition can be thought of as a higher order form of PCA that can be applied to tensors. It decomposes a tensor, X, into a core tensor, G, and a series of matrices referred to as factor matrices, which are orthogonal. In the case of a three-way (three dimensions) tensor, X will have dimensions (I x J x K), G will have dimensions (P x Q x R), and there will be three factor matrices (one for each mode of the tensor): A with dimensions (I x P), B with dimensions (J x Q), and C with dimensions (K x R). P, Q, and R are equivalent to the principal components along each mode. This can be represented by an n-mode vector product of the core tensor and matrices, which is equivalent to the sum of each element of G multiplied by the outer product of each column of each of the matrices A-C, and is shown in Equation 2.33 (183).

$$X \approx G \times_1 A \times_2 B \times_3 C = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_p \circ b_q \circ c_r \qquad (2.33)$$

The $\times_1$, $\times_2$, and $\times_3$, operators are the n-mode vector products, which is when a tensor is unfolded along the nth dimension into a 2D matrix and is multiplied by a matrix. The $\circ$ operator is the outer product of two vectors, which is when the elements of two vectors are multiplied (218, 219). The Tucker decomposition can be generalised to an N-way tensor with N dimensions, according to Equation 2.34 (183).

$$X \approx G \times_1 A^1 \times_2 A^2 \ldots \times_N A^N \quad (2.34)$$

The Tucker decomposition can be calculated using higher order SVD (HOSVD), which is an extension of matrix SVD allowing it to be applied to tensors with more than two dimensions. This method involves using SVD to determine the number of components which accurately explain a mode of the tensor independently of the other modes (183).

The rank of the core tensor, G, can be reduced by reducing the size of the factor matrices (e.g. by reducing the values of P, Q, and R in Equation 2.33). Using reduced factor matrices in the decomposition is referred to as truncated HOSVD. However, truncated HOSVD does not provide the best estimate of the core tensor. This can be improved using higher order orthogonal iteration (HOOI) where an initial Tucker decomposition is calculated using HOSVD and then improved and optimised using an alternating least squares (ALS) algorithm, which iterates until either the error between the core tensor and original tensor does not improve, or until a set number of iterations have passed (183).

The Tucker decomposition can be calculated in Matlab using the tensorlab and N-way toolboxes (220, 221). Tensorlab has an mlrankest function which can be used to estimate the optimal number of components to reduce the rank of the tensor, using HOOI. It varies the number of components over a range of values and plots the error between the core tensor and the original tensor against the compression ratio of the core tensor (222). The elbow of the curve is then selected as the optimal number of components. The Tucker decomposition can then be calculated using the Tucker3 model in the N-way toolbox, which was developed by Claus Andersson and Rasmus Bro to be a fast algorithm which can be used in Matlab (221, 223).

Rank reduction using the Tucker decomposition has multiple application on the literature. Vasilescu and Terzopoulos applied the Tucker decomposition as a dimensionality reduction method to process images of faces used in facial recognition (224), whilst Acar et al. applied it as a dimensionality reduction method to chatroom communication data prior to analysis (225). Muti et al. have applied the Tucker decomposition to filtering methods, using it to extend the Wiener filter so that it can be applied to multidimensional data (226). The Tucker decomposition has also been applied as a denoising method to hyperspectral images by both Bai et al. and Meng et al. (227, 228). In MRI it has been used by Yaman et al. as a dimensionality reduction method to improve the processing of multidimensional MRI data (229); Vargas-Cardona et al. have applied it as a method to interpolate diffusion MRI data, showing that it outperformed standard methods (230); and Brender et al. have applied the Tucker decomposition to denoise MR spectroscopy data (216).

### 2.3.2. The Wavelet Transform

A wavelet is a mathematical function which is used to split up a signal into its constituent frequency components, and then analyse the components according to their scale. To carry out wavelet analysis a 'mother wavelet' or analysing wavelet is defined. The analysing wavelet is then dilated and transformed to form a wavelet basis or wavelet family (231). This basis of functions is defined by equation 2.35. In this equation j and k are integer values related to the dilations and translations of the analysing wavelet respectively, ψ is the wavelet function, and t is time. The wavelets which make up the wavelet basis must satisfy two conditions as shown in equations 2.36 and 2.37. Equation 2.36 suggests that the wavelets must decay and equation 2.37 suggests the wavelets must oscillate (232).

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi\left(2^{-j}t - k\right) \tag{2.35}$$

$$\lim_{t\to\infty}\left|\psi_{i,j}(t)\right| = 0 \tag{2.36}$$

$$\int_{-\infty}^{\infty} \psi_{j,k}(t)dt = 0 \tag{2.37}$$

The discrete wavelet transform can be used to decompose a signal, s(t), into a series of wavelet functions, ψ(t), and wavelet coefficients, a, according to equation 2.38. The coefficients $a_{j.k}$ can be calculated using equation 2.39 (232).

$$s(t) = \sum_{k}\sum_{j} a_{j,k}\psi_{j,k}(t) \tag{2.38}$$

$$a_{j,k} = \left\langle s(t)\psi_{j,k}(t)\right\rangle = \int_{-\infty}^{+\infty} s(t)\psi_{j,k}(t)dt \tag{2.39}$$

The coefficients produced in Equation 2.39 can be split into approximation and detail coefficients using a band pass filter, with the approximation coefficients separated out by a low-pass filter and the detail coefficients separated out by a high-pass filter. The discrete wavelet transform can be repeated over multiple levels, producing a multi-level decomposition as shown in Figure 2.21. Each level of the diagram represents a specific dilation of the analysing wavelet j, and the translations within a level correspond to k (232).

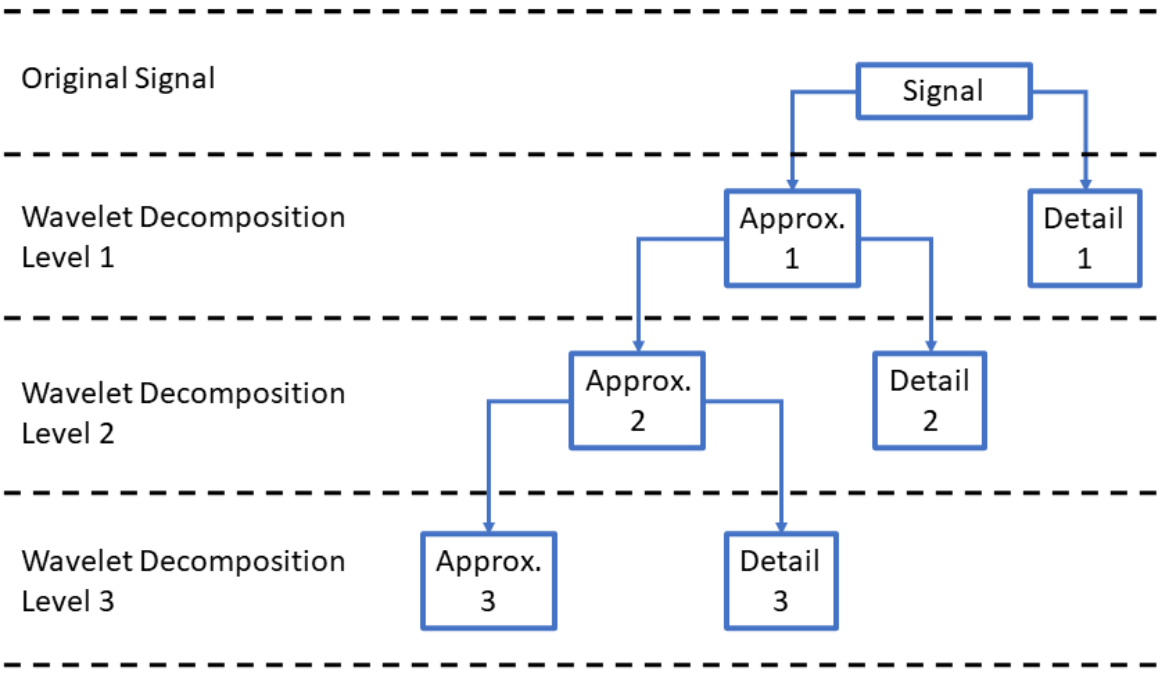



*Figure 2.21: An example of the wavelet decomposition of a signal over several levels.*

#### 2.3.2.1. Wavelet Denoising

Wavelet denoising makes use of the discrete wavelet transform to denoise a signal. When an appropriate wavelet is selected which matches the signal which is being denoised, the high valued coefficients of the decomposition relate to the signal, whilst the lower values coefficients relate to the noise. This is because the wavelet correlates with the signal and not the noise. Therefore, if the low valued coefficients are thresholded out, then the signal can be denoised (232). There are multiple possible methods for determining the thresholds, but a commonly used method is an empirical Bayes threshold (233). The number of levels the signal is decomposed across by the discrete wavelet transform is decided by the user, along with the type of wavelet used in the decomposition (232).Wavelet denoising has been applied in the literature to denoise images and other signals. For example Singh has applied it to denoise natural images (234), Chang et al. applied an adaptive method of wavelet denoising to image denoising (235), and Bnou et al. applied a wavelet denoising approach based on unsupervised learning to image denoising (236). In terms of other signals, Dautov et al. have

applied wavelet denoising to recorded speech (237), Tikkanen has applied it to electrocardiogram signals (238), and Messer et al. have applied it to phonocardiograms (239). Examples of wavelet denoising being applied to MRI data include Cancino-de-Greiff et al. who have applied it to MRI of the liver (232), Ali et al. who have applied it to MR spectroscopy (240), and Wink et al. who have applied it to functional MRI (241).

### 2.3.3 Clustering

Clustering is a form of unsupervised machine learning which is used to group data points into a user defined number of groups. Clustering methods can either be hierarchical or partitional. Hierarchical methods recursively find nested clusters, whilst partitional clustering methods simultaneously find all the clusters (242).

K-means clustering is an example of a partitional method. In k-means clustering, a set of N data points, $X = \{x_1, x_2, \dots, x_N\}$, are separated into K clusters, $C = \{c_1, c_2, \dots, c_k\}$, with cluster means, $\mathrm{M} = \{\mu_1, \mu_2, \dots, \mu_k\}$. The squared error, J, between the points in a cluster and its corresponding cluster mean is given by equation 2.40. The squared error can be summed over all clusters using equation 2.41. The overall aim of the k-means clustering is to minimise this overall squared error (242).

$$J(c_k) = \sum_{x_i \epsilon c_k} \|x_i - \mu_k\|^2 \tag{2.40}$$

$$J(c_k) = \sum_{k=1}^{K} \sum_{x_i \epsilon c_k} \|x_i - \mu_k\|^2 \tag{2.41}$$

Initially, the cluster means are initialised with random values. Then each data point is assigned to the cluster of the cluster mean it is closest to. The cluster means are then recalculated using the newly assigned data points. This process is then repeated through several iterations, until the locations of the cluster means do not move (243).

K means clustering has applications in the literature, for clustering many different types of data. For example Chen et al have applied it to segmenting credit card data (244), Khanmohammadi et al have applied it to clustering medical datasets of patient information to uncover their underlying structure, Dhanachnadra et al have applied it to segmenting images (245), and Wu et al have applied it to brain tumour segmentation in MR images (246).

Hierarchical clustering takes a slightly different approach, in that each individual data point is initially considered to be its own cluster. Each individual cluster is then merged with its nearest neighbour, and this process is repeated until the desired number of clusters are obtained (247). In hierarchical clustering two factors can be varied: the clustering method, which defines the data points used to measure the distance between two clusters; and the distance measure, which defines how that distance is calculated. Clustering methods which may be used include: the nearest neighbour, furthest neighbour, or a group average (which may be weighted or unweighted) (248).

Distance measures which may be used include: the Euclidean distance (the direct distance between two data points), the city-block distance (the sum of the difference in each coordinate between the two data points) and the Chebyshev distance (the maximum difference value between two coordinates). Of these, the Euclidean distance is most commonly used (248).

Hierarchical clustering has multiple applications in the literature. For example, Byron et al. have applied it to a dataset of student background and performance in order to understand its underlying structure (249), Govender and Sivakumar have reviewed its use in analysing air pollution data (250), and Arifin et al. have combined it with histogram thresholding to produce an image segmentation method (251). Examples of it being applied in MRI include Cordes et al. and Filzmoser et al. who used it to analyse functional MRI data (252, 253), and

Vupputuri et al. who used it to automatically segment ischemic regions from MRI scans of stroke patients (254).

### 2.3.4. Classification

Classification is a form of supervised machine learning, in which a learner is provided with a labelled dataset consisting of pairs of input variables and a corresponding output variable. The learner then uses this dataset to predict the output variables of unlabelled data, consisting of just the input variables. The process of classification is summarised in Figure 2.22 (181). The following subsections cover: validation, the machine learning algorithms used in this work, methods to optimise the classifiers, and methods to assess the performance of classifiers.

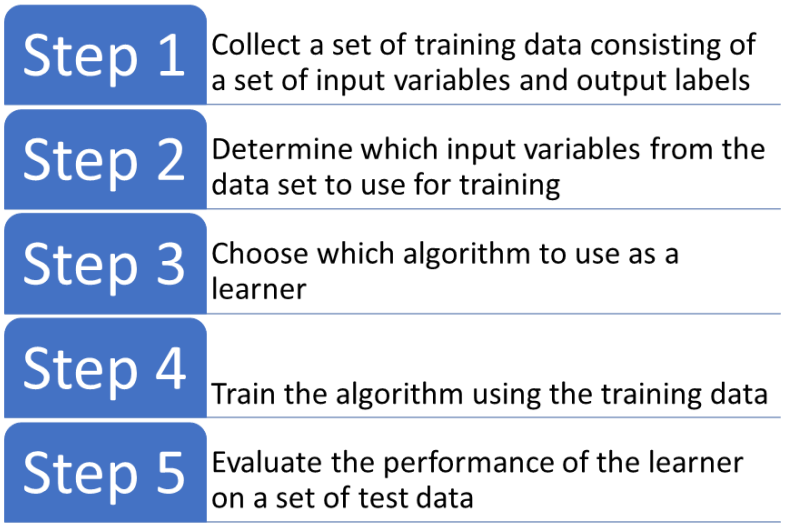



*Figure 2.22: Steps involved in machine learning classification.*

Classification can also be carried out by deep learning, which uses neural networks. Neural networks are built using artificial neurons, which take the weighted sum of a number of inputs and then apply an activation function to the result. The weights of each artificial neuron are adjusted during training (255). The resulting network has many applications, for example they can be used for classification, image processing, speech recognition and natural language processing (256). Neural networks differ from other classification algorithms is that classification algorithms require metrics to be extracted from the data before the machine

learning algorithm can be applied, whilst neural networks process the raw data (257). However, one of the downsides is that they require a lot of training data. As the datasets used for machine learning in this work were of limited size, neural networks were not considered in this work.

#### 2.3.4.1. Validation

Validation is an important step in assessing the performance of a classifier. If the same data is used to train and test the classifier then there is a risk of the classifier over-fitting to the data. Therefore, it's necessary to split the data into training and testing data, and the three most commonly used methods to split the data are: resubstitution, holdout and k fold cross-validation (258).

In resubstitution all the data is used for training and testing. In holdout the data set is split into training and testing data (often the training data set is larger). In k fold cross-validation the data set is split into k equally sized 'folds', the classifier is then trained on (k-1) folds and tested on the remaining fold. This process is repeated until all the folds have been used as the testing data, and the performance metrics are averaged across all the folds (258). A slight variation on k fold cross-validation is stratified cross-validation, where it is ensured that each fold has the same distribution of data (259).

These methods have been previously compared in the work by Kohavi et al. and stratified k fold cross-validation was shown to be the preferred method as it is less biased compared to the other validation methods (259). They recommended that validation should be carried out using stratified cross-validation with 10 folds (259).

#### 2.3.4.2. Algorithms

One of the crucial steps is the selection of the learning algorithm. There are multiple different learning algorithms that can be applied and examples include: decision trees, support vector

machines, ensembles, random forests and logistic regression algorithms (181). The following sections provide an overview of how each algorithm works and gives examples of the algorithms being applied in the literature. This is not an exhaustive list of machine learning algorithms, and those described below are the ones applied in this thesis. They were selected as they were suited to the datasets used in this thesis.

#### 2.3.4.2.1. Binary Tree

Decision trees are algorithms which generalise a dataset by deriving a set of rules, which is represented by a tree. The decision tree consists of decision nodes, leaves, and branches. Decision nodes are used to split the data based on a test; leaves are the endpoint of the tree, where the class labels are assigned; and branches connect the leaves and the nodes. The node at the top of the tree with no branches input into it is the root node. At each node there is a rule which is used to split the data, if the decision at each node is either true or false, then the decision tree is a binary tree (260).

Binary decision tree algorithms build a decision tree by splitting the data at each node using a condition. In order to decide how to split the data and whether to split the data, a 'gain' measure is defined which is used to quantify any improvement from the split (261). Gain is calculated by calculating the 'impurity' of the node prior to the split and the weighted average of the impurities of the two nodes created by the split (262). There are several different methods to define the gain and studies have shown that the choice of method has little effect on the outcome, as many methods are consistent with each other (263).

An important part in growing a binary tree classifier is deciding when to stop splitting the data. It is possible to continually split the data until each node has one datapoint in it, but this would result in overfitting the data. The three main methods for stopping are: defining a maximum number of splits in the tree, defining a maximum number of datapoints in a leaf,

and defining a minimum value for the gain of a split (262). Figure 2.23 summarises the process of training a binary tree classifier.

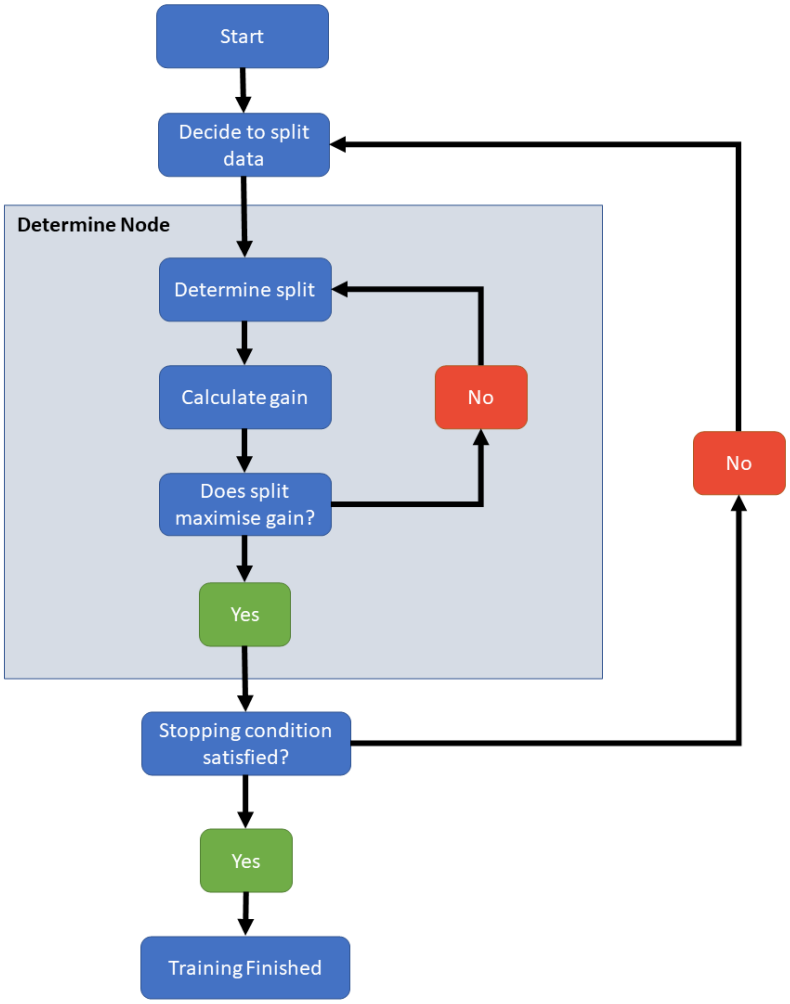



*Figure 2.23: A flow chart describing the process of training a binary tree classifier.*

The binary tree classifier has been applied to classification of MR images in multiple applications in the literature. For example, Al-Badarneh et al. have used a binary tree classifier applied to features extracted from MRI images to classify brain as 'normal' or 'abnormal' (264), whilst Fayaz et al. applied a binary tree to features extracted from histograms of MR images to also classify them as 'normal' or 'abnormal' (265).

#### 2.3.4.2.2. Support Vector Machine (SVM)

An SVM classifier splits a dataset into two classes using a hyperplane (266). The points closest to the hyperplane are referred to as support vectors as they are the most difficult to

classify, and their location has the largest effect on the position of the hyperplane. The separation between the hyperplane and the support vectors is defined as the margin, and the best fitting SVM model is the one which produces the largest margin (267). For an SVM the input data can be considered to be a 2 dimensional matrix, X, with n rows equal to the number of observations, and p columns equal to the number of predictors, whilst the classification labels can be considered to be a vector, Y, with n rows, where each value is 1 or -1. X can be represented by a set of vectors, $x_i$, (i = 1, …, n) where each vector is a row of X, whilst the elements of Y can be represented by, $y_i$ (i = 1, …, n). The hyperplane, $f(x)$, is represented by Equation 2.42, where $\omega^T$ is the transpose of an n-dimensional vector and b is a bias term (266).

$$f(x) = \omega^T x + b = 0 \tag{2.42}$$

The hyperplane which provides the best separation is found by finding ω and b to minimise the value of ||ω|| and produce a hyperplane which satisfies Equation 2.43. As the support vectors define the boundary of the classifier they will satisfy Equation 2.44 (267).

$$y_i f(x_i) \geq 1 \tag{2.43}$$

$$y_i f(x_i) = 1 \tag{2.44}$$

If the data is not linearly separable then a soft margin can be used, which means that the hyperplane separates the majority of but not all datapoints. However, in some cases the data may not be separable even with a soft margin. In this case a kernel function is used to transform the data into a higher dimensional space where it may be easier to separate with a hyperplane (267). There are multiple types of kernel function which can be used to transform the data, but the three most commonly used are the Gaussian kernel, the linear kernel, and the polynomial kernel, which are represented in equations 2.45 to 2.47, respectively. In all three

equations $x_i$ and $x_j$ represent two different observations from X. In equation 2.45, σ is a constant which can be altered by the user. In equation 2.47, p is the order of the polynomial (266).

$$\mathrm{K}(x_i, x_j) = e^{\frac{\|x_i - x_j\|^2}{2\sigma^2}} \tag{2.45}$$

$$\mathrm{K}(x_i, x_j) = x_i^T x_j \tag{2.46}$$

$$\mathrm{K}(x_i, x_j) = \left(x_i^T x_j + 1\right)^p \tag{2.47}$$

Examples of SVM classifiers applied to MRI include: work by Gupta et al. where a SVM classifier was used to identify slices of MRI datasets which contained tumour (268); work by Sujitha et al. which used an SVM classifier applied to features extracted from MR images to predict the presence of epilepsy (269); and work by Gonella et al. used pixel intensities combined with features extracted from MR images to segments brain tumours (270).

#### 2.3.4.2.3. Ensemble

Ensemble classification combines classifiers to produce improved performance. It uses classifiers defined as 'weak learners' (classifiers with a classification error similar to random guessing) and combines them to produce a 'strong learner' (a classifier with a low classification error). There are multiple different methods for combining weak learners (271). A common method of combining classifiers is boosting, and boosting algorithms which use a binary classification tree as the weak learner have shown to consistently improve performance, when compared to a single binary classification tree (272).

Adaptive boosting (also referred to as AdaBoost) uses the weighted sum of a set of weak learners to produce a final output. A series of weak learners are applied iteratively to a weighted training dataset consisting of N data points. The weights are then updated using the

results from the weak learner. The process is repeated over M iterations. Initially, the weight on each datapoint within the training data is set to 1/N. For the mth iteration the weighted error of the weak learner, $err_m$, and the weight of the weak learner, $\alpha_m$, are calculated using equations 2.48 and 2.49. In equation 2.48, $w_i^m$ is the weights from the mth iteration, $y_i$ is the expected class label, $H_m(x_i)$ is the predicted class label by the weak learner, and h() is the Heaviside function, which is equal to 1 if $-y_i H_m(x_i) \geq 0$ and equal to zero if $-y_i H_m(x_i) < 0$. The new weights, $v_i^m$, are then calculated using equation 2.50, and are then normalised (271).

$$err_m = \sum_{i=1}^{N} w_i^m h\left(-y_i H_m(x_i)\right) \tag{2.48}$$

$$\alpha_m = \frac{1}{2}\log\left(\frac{1 - err_m}{err_m}\right) \tag{2.49}$$

$$v_i^m = w_i^m \exp\left(-\alpha_m y_i H_m(x_i)\right) \tag{2.50}$$

A final result is determined from the weighted sum of the classification results of each weak learner. The advantage of this method is that the weightings allow subsequent weak learners to focus on the misclassified data (271).

There are several variations on AdaBoost available. Logit Boost and Gentle AdaBoost are two variations which are similar in that they both use a weighted least squares method to train the weak learner during each iteration of the algorithm. The weights used are the ones updated during each iteration of the algorithm. Gentle AdaBoost, trains the weak learner using the expected class labels. Logit Boost calculates a class probability estimate for each data point in the training data. The weights are then updated using the class probabilities, and a working response is calculated using the actual class labels and the class probabilities. The working response values are then used in place of the actual class labels when training the weak

learner. The aim of these two variations is to provide more reliable and stable ensembles (271).

A further variation on AdaBoost is random undersampling boosting (RUSBoost), which introduces random undersampling to the AdaBoost algorithm to help deal with class imbalances, which can lead to biased classifiers. Random undersampling randomly removes datapoints from the majority class until a desired split between classes is achieved. In RUSBoost, random undersampling is applied after the weight have been calculated and before the weak learner is trained on the data (273).

Boosting classifiers have been applied to classify MR images in multiple applications. For example: Minz et al. have applied AdaBoost to determine brain tumour type using features extracted from MRI images (274), and Savio et al. have applied it to features extracted from MR images to detect Alzheimer's (275). Gentle Boost has been applied by Pang et al. to glioma grading based on features extracted from MRI (276), and Zhang et al. have applied it to assess spinal curvature in CT imaging (277). Logit Boost has been applied by Zhang et al. to EEG data to diagnose alcoholism and to MRI data to diagnose ADHD (278); whilst Gitto et al. have applied it to features extracted from CT imaging to classify brain tumours (279). RUSBoost has been applied by Maglietta et al. to segment the hippocampal region of the brain on MR images (280); and Tong et al. have applied it to features extracted from MR images to distinguish between different types of neurodegenerative disease (281).

#### 2.3.4.2.4. Random Forest

A random forest classifier consists of a series of decision trees, which are combined using bootstrap aggerating (bagging). The bagging process is used to improve the stability and accuracy of the classifier. To build a random forest classifier, N datapoints are sampled (with replacement) from the training data, and p features are selected (where p is less than the

number of features and is often $\sqrt{p}$ or p/3). A binary tree is then trained following the process described in Section 2.3.4.2.1. This process is then repeated t times, where t is the number of trees. Therefore, for a random forest classifier the values of N, p, and t can be selected to optimise the model (282).

To make a prediction using a random forest classifier, the result from each of the individual trees is aggregated to produce a final result. In the case of a random forest classifier, the final result is selected by a majority vote of the classes selected by each tree. This process is demonstrated in Figure 2.24 (282).

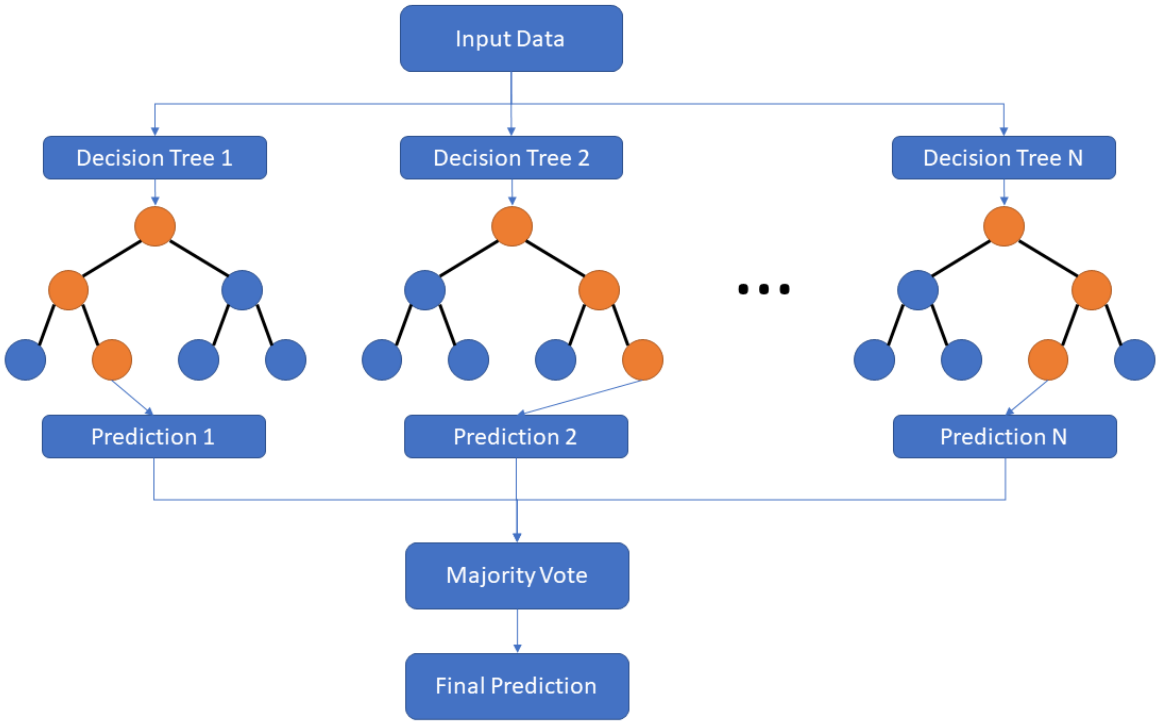



*Figure 2.24: A flow chart demonstrating how a random forest classifier with N trees makes a prediction.*

The random forest classifier has been applied to classification problems in MRI. For example, Payabvash et al. have used it to determine tumour type in patients with posterior fossa tumours using features extracted from diffusion weighted MRI (283); and Yang et al. have applied the random forest to produce pseudo-CT images from MRI data (284).

#### 2.3.4.2.5. Logistic Regression

In logistic regression classification a logistic function is fitted to the data and is used to predict the class. The logistic function is a sigmoid function and takes the form shown in equation 2.51, where σ is the sigmoid function, and t is the input variable (285).

$$\sigma(t) = \frac{1}{1 + e^{-t}} \tag{2.51}$$

The sigmoid function produces values between 0 and 1, so the output of logistic regression classification is the probability that the data belongs to a specific class. Equation 2.51 only works for predicting the probability of an item belonging to one of two classes based on one input variable. If there are multiple input variables then Equation 2.51 needs to be expanded. The variable t can be rewritten for a dataset with d variables according to Equation 2.52, where $x_{i,d}$, are the variables, i, is the number of datapoints, and ω is the parameters of the logistic regression model (285).

$$t = \omega_0 + \omega_1 x_{i,1} + \omega_2 x_{i,2} + \cdots + \omega_d x_{i,d} \tag{2.52}$$

To train a logistic regression model the values of the model parameters, ω, must be determined. To do this a loss function is defined. A typical loss function is cross entropy, $L_i$, shown in equation 2.53, where $p_i$ is the predicted label and $y_i$ is the actual label (286).

$$L_i = -y_i \ln(p_i) - (1 - y_i) \ln(1 - p_i) \tag{2.53}$$

The cross entropy loss function produces large values for incorrect classification and small values for correct classifications. An average loss function can then be calculated for the entire dataset, and the values of the model parameters are varied over a number of iterations to minimise the loss function (286).

Examples of logistic regression being applied to MRI in the literature include: work by McLaren et al. who applied logistic regression to features extracted from DCE-MRI data to distinguish between malignant and benign breast legions (287); whilst Xiao et al. have applied it to features extracted from MRI to diagnose Alzheimer's disease (288).

#### 2.3.4.3. Optimising Classifiers

Parameters are quantities that are estimated by the classifier during training, whilst hyperparameters are quantities of the classifier which are set by the user. For example, a hyperparameter of a binary tree classifier is the maximum number of splits used when constructing the tree. Adjusting the hyperparameters can improve the performance of a classifier, and they can be adjusted manually, or by using an automated process (289). In hyperparameter optimisation an objective function, f(x), is defined where the input of the function, x, is a given combination of hyperparameters, and the output of the function is a measure of the classifier performance. The function f(x) is unknown and is considered to be a 'black-box' function (289).

The simplest method to optimise the hyperparameters is manual adjustment, but it is also inefficient, can be difficult to reproduce, and requires the user to have expert knowledge (289). This process can be automated using the grid search method, where every combination of hyperparameters is tested and the best performing combination is then selected. If one or more of the hyperparameters is continuous then they are split into a set of discrete values. This is quicker than manual searching and removes the need for expert knowledge to optimise the hyperparameters. However, searching through all possible combinations of hyperparameters can be very computationally expensive (290). An alternative is the random search method, where the combinations of hyperparameters are chosen randomly, instead of searching all possible combinations. This has been shown to offer similar performance to grid

searching, whilst being less computationally expensive. However, this is not a universally applicable method, as it does not perform as well with complex machine learning algorithms (291).

An alternative to the searching based methods above is Bayesian optimisation, which has been shown to perform better than standard searching methods (289, 292, 293). The aim of this method is to approximate f(x) from a series of sample points. It uses Bayes theorem to state the relationship between the model of f(x), M, and the evidence from already sampled points of f(x), E, as shown in equation 2.54. In this equation, P(M) is the probability of the model, P(E|M) is the probability of E given M (the likelihood), and P(M|E) is the probability of M given E (the posterior probability) (289).

$$P(M|E) \propto P(E|M)P(M) \tag{2.54}$$

P(M|E) is the posterior distribution of f(x) and is determined from sampled points of f(x) using equation 2.54. The maxima of the posterior distribution is then found using an acquisition function, u. There are several different acquisition functions which can be used. The two most common are the probability of improvement and the expected improvement. Probability of improvement searches near the current optimal value to find points which are likely to become the new optimal value, whilst expected improvement calculates the degree of improvement between the current optimal value and other local values. A large value of u correlates to a large value of f(x), so finding the maxima of u is equivalent to finding a maxima of f(x). Once the maxima of u has been found the corresponding value for x can be used to sample f(x). This new set of values is then added to the samples of f(x) (E in equation 2.51). The new samples are used to recalculate the posterior probability distribution. This process is repeated until either a set number of iterations have passed, or the difference

between the current maxima and the next maxima reaches a minimum value (289). Figure 2.25 summarises the process in a flow chart.

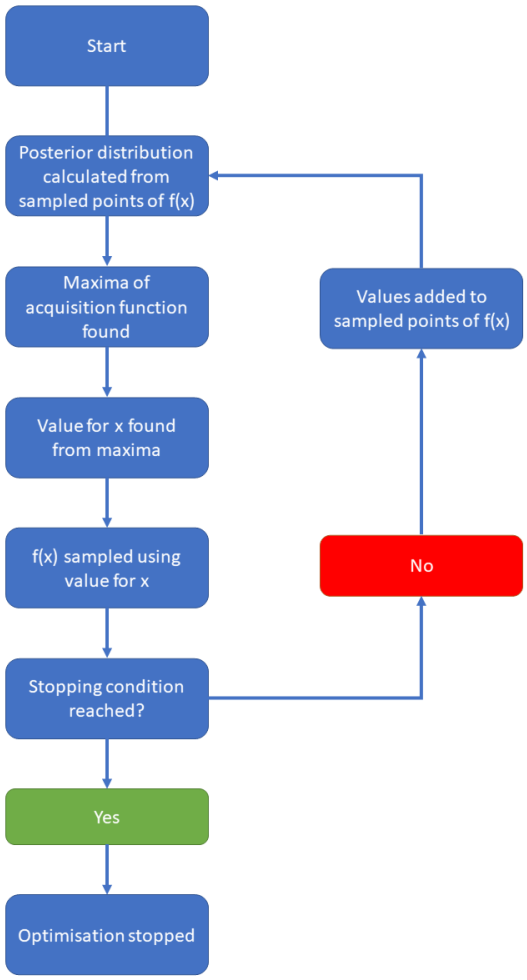



*Figure 2.25: A flowchart summarising the process of Bayesian optimisation.*

### 2.3.4.4. Assessing Classifier Performance

Classifier performance can be assessed using a confusion matrix and the metrics calculated from it. A confusion matrix, shown in Figure 2.26, is a graphical way of comparing the predictions made by the classifier to the actual class labels (294).

| | | True Class | |
|---|---|---|---|
| | | Positive | Negative |
| Predicted Class | Positive | True Positive | False Positive |
| | Negative | False Negative | True Negative |

*Figure 2.26: A confusion matrix used to calculate metrics to assess classifier performance.*

From the confusion matrix a series of performance metrics referred to as sensitivity, specificity, precision, and classification error can be calculated. They are defined by equations 2.55-2.58

(294).

$$sensitivity = \frac{true\ positives}{true\ positives + false\ negatives} \tag{2.55}$$

$$specificity = \frac{true\ negatives}{true\ negatives + false\ positives} \tag{2.56}$$

$$precision = \frac{true\ positives}{true\ postives + false\ positives} \tag{2.57}$$

$$classification\ error = \frac{false\ positives + false\ negatives}{total\ data\ points} \tag{2.58}$$

Another method of assessing classifier performance is a receiver operating characteristic (ROC) curve, which is a plot of the true positive rate (sensitivity) against the true negative rate (1-specificity), for a range of thresholds used to separate the classes. The area under the curve (AUC) is a measure of the performance of the classifier, with a value of 1 representing a perfect classifier (294). Figure 2.27 shows an example ROC curve.

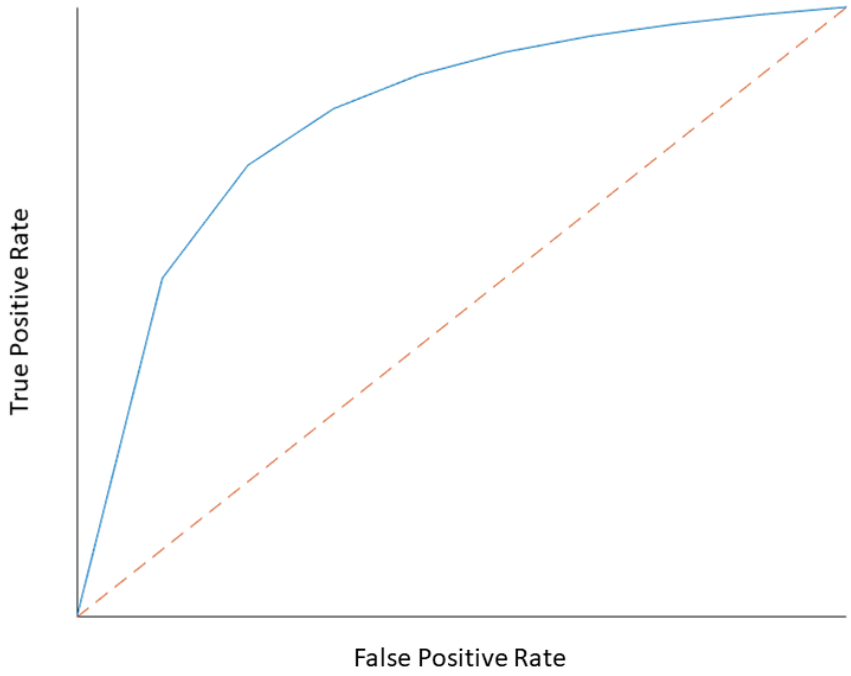



*Figure 2.27: An example of an ROC curve*

### 2.3.5. Justifications

The computational techniques described in this section will be applied to DSC-MRI datasets. A DSC-MRI dataset consists of a series of images recorded over time, with one image recorded for each slice and time point. Therefore, the entire dataset can be represented as a 4D tensor, with the first two dimensions representing the rows and columns of the image, the third dimension representing the slices, and the fourth dimension representing the signal-time courses. Similarly, a single slice can be represented by a 3 dimensional tensor with two dimensions of the image and the third dimension for the signal-time courses.

For the machine learning classifiers, the training and testing data consisted of four features calculated from the signal-time courses, which are SDNR, RMSE, FWHM, and PSR. The target variables were the results of carrying out qualitative review on the signal-time courses. Table 2.2 addresses each technique and explains why it is suited to the data described above.

| Technique | Reason why it's well suited to the data in the thesis | Reference |
|---|---|---|
| SVD | SVD can be used to decompose a matrix into component matrices. MRI data can be represented as a matrix. | (185) |
| PCA | PCA uses SVD to reduce the dimensionality of a dataset. It can be applied to matrices, and DSC-MRI data can be represented as a matrix. | (198) |
| Matricization | DSC-MRI data can be represented as a tensor with more than two dimensions, to include all time points, all slices, or both. Matricization can be used to unfold this into a matrix, which allows SVD and PCA to be applied to it. | (216) |
| Tucker Decomposition | Tucker decomposition is well suited to highly dimensional data and DSC-MRI data is highly dimensional. Tucker decomposition allows the data to be represented in a more compact way which allows it to denoise more effectively than SVD. | (216) |
| Wavelets | Wavelets are well suited to approximate data with sharp discontinuities. The signal drop in a DSC-MRI signal-time course is a relatively sharp discontinuity. | (231) |
| K-Means Clustering | It's a simple, efficient, and widely used method of clustering data where the underlying structure is unknown | (242) |
| Hierarchical Clustering | Hierarchical clustering is a simple and reproducible clustering technique which has been applied across many data sets. | (295) |
| Binary Tree | The binary tree is a simple and efficient classification method, which has been shown to be more robust to outliers than other methods. | (296) |
| Support Vector Machine | The support vector machine performs well when there are a small set of features. The dataset used in this thesis has four features | (266) |
| AdaBoost | Boosting allows classifiers to be trained with less bias and better performance than a single classifier and AdaBoost is suited to small datasets. | (271, 297) |
| Logit Boost | Logit Boost has the same benefits as AdaBoost but can produce more stable results. | (297) |
| Gentle AdaBoost | Gentle AdaBoost has the same benefits as AdaBoost but can produce more stable results. | (297) |
| RUSBoost | RUSBoost has the same benefits as AdaBoost but is also well suited where there is a class imbalance in the training data. The dataset used in this thesis has a class imbalance. | (273, 297) |
| Random Forrest | Bagging produces more stable classifiers than an individual classifier, with less bias and better performance than a single classifier. It is also suited to small datasets. | (271, 297) |
| Logistic Regression | Simple and efficient model which is well suited to binary classification problems | (298) |

*Table 2.2: A summary of the reasons why each computational technique is suited to the data used in this thesis.*

## 2.4. Statistical Methods

This section covers the theory behind some of the statistical methods used throughout the thesis.

### 2.4.1. Cohen's Kappa

Cohen's kappa, κ, is a statistical measure for assessing the reliability between two raters and was developed by Jacob Cohen in the 1960's (299). Cohen's kappa can be calculated using equation 2.59, where $Pr(a)$ is the actual observed percentage agreement and $Pr(e)$ is the expected (chance) agreement (300).

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(\mathrm{a})} \tag{2.59}$$

Pr(e) is calculated from equation 2.60 and Table 2.3 summarises the chi-square table used to calculate the parameters within equation 2.60 (300).

$$\Pr(e) = \frac{\left(\frac{cm^1 \times rm^1}{N}\right) + \left(\frac{cm^2 \times rm^2}{N}\right)}{N} \tag{2.60}$$

| | | Rater 1 | | Row Marginals |
|---|---|---|---|---|
| | | Normal | Abnormal | |
| Rater 2 | Normal | A | B | $rm^1$ = A + B |
| | Abnormal | C | D | $rm^2$ = C + D |
| Column Marginals | | $cm^1$ = A + C | $cm^2$ = B + D | N = $rm^1$ + $rm^2$ + $cm^1$ + $cm^2$ |

*Table 2.3: A summary of the parameters calculated from the chi square table, used in the calculation of Pr(e).*

The Cohen's kappa statistic can have values ranging from 0 to 1 and there are a set of pre-determined cut-offs that define the level of agreement, as summarised in Table 2.4 (300).

| Value of Kappa | Level of Agreement |
|---|---|
| 0 – 0.2 | None |
| 0.21 – 0.39 | Minimal |
| 0.40 – 0.59 | Weak |
| 0.60 – 0.79 | Moderate |
| 0.80 – 0.90 | Strong |

| | |
|---|---|
| > 0.90 | Almost Perfect |

*Table 2.4: Cut-offs for the Kappa statistic.*

### 2.4.2. Comparing Means

Statistical testing can be used to find significant differences between two sets of values. The type of test used depends on whether the distribution of values that are being compared is normal or non-normal. The normality of a set of values can be tested using the Shapiro-wilk test for normality. This then determines the type of statistical test used (301).

#### 2.4.2.1. Normal Distribution

For a normally distributed set of values, the Analysis of Variance (ANOVA) technique can be used. This test is based on the variance in the means of the sets of values. For each set of values, a group mean and several sample means are calculated. An overall grand mean (average of all the group means) is also calculated. The mean square between groups, $MS_B$, and within groups, $MS_W$, is then calculated, and is given by equations 2.61 and 2.62, respectively. $SS_B$ is the sum of squared deviations of the sample means compared to the grand mean, $SS_W$ is the sum of squared deviations of the sample means around the group mean, $df_B$ is the degrees of freedom between groups (given by the number of groups minus 1) and $df_w$ is the degree of freedom within a group (given by the number of values minus 1) (302).

$$MS_B = \frac{SS_B}{df_B} \tag{2.61}$$

$$MS_W = \frac{SS_W}{df_W} \tag{2.62}$$

Dividing $MS_B$ by $MS_W$ produces an F statistic which gives a measure of the difference between the groups. For this test, the null hypothesis states that the means are equal. Since the F statistic belongs to an F distribution the P value of the null hypothesis is equal to the P value from the F distribution (302).

#### 2.4.2.2. Non-Normal Distribution

The Kruskal-Wallis test can be used to compare means when the values do not follow a normal distribution. The first stage is to rank all the values within the dataset (ignoring which group they belong to) by value, where the lowest value is ranked 1 and the largest value has a rank equal to the number of values within the dataset. Repeated values are assigned average ranks. The test statistic $H$ is then calculated, according to equation 2.63, where $R_i$ is the sum of the ranks within the ith group, $n_i$ is the number of data points in the ith group and $N$ is the total number of data points within the dataset (303).

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - e(N+1) \tag{2.63}$$

$H$ is a value representing the variance in the ranks between groups with a correction applied to account for any average ranks. For this test, the null hypothesis is that the means of the groups are equal. Since $H$ follows a chi square distribution the probability of the null hypothesis is equal to the P value if $H$ corresponds to a chi-square value (304).

## 2.4. Summary

This chapter covers the theory behind MRI, DSC-MRI and the computational and statistical methods used in this thesis. The MRI theory and DSC-MRI theory provide the background theory needed to understand the work carried out in the thesis, and the DSC-MRI section outlines some of the issues relating to assessing data quality, which will be addressed in the results chapters.

The computational methods section covers the theory behind the different techniques used, examples of them being applied in the literature, and justifications for why they are suited to this data. Combining the different techniques is useful as it combines the advantages of each

technique and allows an image processing pipeline to be produced, which is one of the aims of the thesis.

The statistical methods section covers the theory behind the standard statistical methods used in this work. Multiple statistical tests are required as the test used depends on the distribution of the data it is applied to.

# Chapter 3: An Automated Machine Learning Method To Assess the Quality of DSC-MRI Data

# 3. An Automated Machine Learning Method to Assess the Quality of DSC-MRI data

## 3.1. Introduction

DSC-MRI is prone to motion and susceptibility artefacts, which degrade the quality of acquired data (34). The scanner and acquisition protocol for DSC-MRI also commonly varies from centre-to-centre, which affects the signal-time courses produced and the SNR of the CBV maps (305, 306), limiting the clinical applicability of the technique (61). Currently, the American Society of Functional Neuroradiology (ASFNR) recommendation for quality control of DSC-MRI data is to assess signal-time courses by eye, using qualitative review (34). However, there can be discordance between reviewers and one DSC dataset contains thousands of signal-time courses, so it is not practical to assess the quality of all signal-time courses. An automated process, which reflects the qualitative review process and could be applied to all signal-time courses, would be much more applicable.

Previous work has shown that it is possible to define statistical thresholds and apply these to quality metrics calculated from DSC-MRI signal-time courses to automatically assess data quality (307). This used three metrics (full width half minimum (FWHMin), percentage signal recovery (PSR) and % failed gamma variate fit rate) averaged for each dataset. However, using average metrics to assess quality may not give a true reflection of data quality. Machine learning has also previously been applied to DSC-MRI data for several different applications, but not for assessing data quality. For example, it has been used to generate perfusion maps from DSC-MRI data by McKinley et al. (308), predict survival in glioma patients by Emblem et al. (309), and classify tumour type using DSC-MRI alone by Jeong et al. and Ji et al. (310, 311) and in combination with other MR modalities, such as diffusion weighed imaging (DWI), diffusion tensor imaging (DTI), magnetic resonance spectroscopy (MRS), and DCE-MRI by Grist et al., Hu et al., Tsolaki et al. and Swinburne et al. (39, 312-314).

The studies presented above use a variety of features as predictors. Most studies use metrics calculated from the DSC-MRI data but the work by McKinley et al. uses time points from the concentration time curves as predictors (308). However, this requires all the signal-time courses to be interpolated as a pre-processing step, which may produce misleading results when trying to assess signal-time course quality, as artefacts may be obscured. Some studies use features from DSC-MRI in combination with features from other modalities. However, this would not be practical in assessing DSC-MRI data quality as data quality will not be dependent on features extracted from other MR imaging modalities. Based on the literature above, it may be possible to apply machine learning to features extracted from DSC-MRI signal-time courses to determine data quality.

The overall aim of this chapter is to develop an objective method to assess whether a DSC-MRI dataset acquired in the paediatric brain has sufficient quality for clinical use. To achieve this, the objectives of this chapter are to: assess the discordance in qualitative review between two reviewers; use qualitative review to determine a set of thresholds in quantitative measures of quality; and investigate if an automated quality control aid can be devised using the quantitative measures of quality and qualitative review results.

## 3.2. Methods

### 3.2.1. Patient Data

The patient data is from an imaging study entitled “CNS 2004 10 Functional Imaging of Tumours” (NRES REC ref: 04/MRE04/41). This is a multi-centre study sponsored by the University of Birmingham ref: RG_09-028 and UoB Ethics ref: ERN_11-1170. Informed consent was obtained from all patients included in the study.

For this study, a dataset containing 25 paediatric patients, from four different centres (Birmingham, Liverpool, Newcastle, and Nottingham), was used. The acquisition protocols

are summarised in Table 3.1. Patients from centres 1-3 were all injected with Dotarem contrast agent (Dotarem, Guerbert, France), whilst patients from centre 4 were injected using either Gadovist (Bayer AG, Leverkusen, Germany) or Magnevist (Bayer AG, Leverkusen, Germany). All patients received a total dosage of 0.1 mmol/kg, using a power injector. Patients who did not receive a pre-bolus were injected using a single bolus protocol and received the total dosage in one injection during the DSC acquisition. Patients who received a pre-bolus were injected using a split bolus protocol and received half the total dosage prior to the DSC acquisition, and the other half of the dosage during the DSC acquisition. This was to minimise $T_1$ weighted effects from contrast agent leakage (as discussed in section 2.2.2.2.). For all patients, a volume of up to 10 ml of saline was injected after the total dosage of contrast agent was received. All injections were administered at a rate of 3 ml $s^{-1}$.

| Centre | No. of Patients | No. of Signal-Time Courses | Field Strength (T) | Pre-Bolus | Sequence | Flip Angle (°) | TE (ms) | TR (ms) | Voxel Size (mm) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 139 | 3 | yes | GE-EPI | 20 | 40 | 1829 - 4865 | 2.5 x 2.5 x 3.5 |
| 1 | 5 | 193 | 1.5 | yes | GE-EPI | 20 | 40 | 1490 - 1643 | 2.4 x 2.4 x 5 |
| 1 | 3 | 112 | 3 | no | sPRESTO | 7 | 22 | 15 | 3.4 x 3.4 x 3.5 |
| 2 | 5 | 177 | 3 | yes | GE-EPI | 75 | 40 | 1335 - 2343 | 1.75 x 1.75 x 4 |
| 3 | 3 | 166 | 3 | no | GE-EPI | 45 | 29 | 1570 | 3.4 x 3.4 x 3.5 |
| 4 | 3 | 112 | 3 | yes | GE-EPI | 20 | 40 | 1865 | 2.5 x 2.5 x 3.5 |
| 4 | 2 | 96 | 1.5 | no | sPRESTO | 7 | 25 | 17 | 3.4 x 3.4 x 3.5 |
| 4 | 1 | 16 | 3 | no | sPRESTO | 7 | 24 | 16 | 1.8 x 1.8 x 3.5 |

*Table 3.1: A summary of the acquisition protocols used to acquire the patient data, and the number of patients and signal-time courses from each acquisition protocol.*

### 3.2.2. Qualitative Review of Patient Data

A total of 1,027 signal-time courses were used in the qualitative review process, which were extracted from 25 patients. A large number of patients were used to ensure a range of artefacts and acquisition protocols were included, and Table 3.1 summarises how many signal-time courses came from each acquisition protocol. Signal-time courses were randomly selected from a set of manually pre-defined regions within each patient, which included: grey matter (GM), white matter (WM), the edge of the brain, the edge of the ventricles, and the cerebellum. Signal-time courses were selected from slices that did not contain tumour, by selecting supratentorial slices from patients with infratentorial tumours and infratentorial slices from patients with supratentorial tumours. All signal-time courses were assessed by qualitative review by Stephen Powell (a PhD student with 3 years experience), and a randomly selected subset of 243 signal-time courses were additionally assessed by Dr Stephanie Withey (a clinical scientist specialising in MRI with 8 years experience). Comparing the differences in qualitative review between two reviewers can be used to estimate the discordance between reviewers. Qualitative review was carried out using the guidance from the ASFNR recommendations (34). This involved assessing whether a clear signal drop was present and the level of noise within the signal. For the subset of signal-time courses assessed by two reviewers the qualitative review results from author 2 were used as the ground truth.

The qualitative review results for the 784 signal-time courses assessed only by author 1 and the 243 signal-time courses assessed by author 2 were combined to form a ‘quality control test dataset’.

### 3.2.3. Calculating Quantitative Measures of Quality

The first step in automating the quality control process was calculating the metrics which were used as quantitative measures of quality, for all the signal-time courses within the

quality control test dataset. This included: signal drop to noise ratio (SDNR), root mean square error (RMSE), full width half maximum (FWHM) and percentage signal recovery (PSR). These measures were calculated using equations 2.16 – 2.18 shown in section 2.2.3.1 of the theory chapter. The RMSE values were normalised to the area of the first pass to allow for comparisons between patients and centres.

To calculate the quantitative measures, it was necessary to define the dynamic time point at which the baseline ended, and the post-bolus acquisition started. The end of the baseline was determined by calculating the moving mean (with sliding window of three) and cumulative mean of the signal-time course, starting from the first dynamic, and finding the point where the means diverged. The start of the post-bolus signal was determined using the same process but starting from the last dynamic.

All signal-time courses were then ordered by ascending SDNR and the range of SDNR values where signal-time courses pass and fail quality control was investigated. This process was then repeated for each of the quantitative measures.

### 3.2.4. Thresholds from Qualitative Review

Threshold values for each quantitative measure were determined using the qualitative review results from the quality control test dataset, and k-fold cross validation (CV) with k = 10. In k-fold CV the data is separated into k equally sized folds. (k-1) folds are used as training data and the remaining fold is used as testing data, from which the performance metrics are calculated. This process is repeated until all folds have been used as testing data and average performance metrics are calculated (315). For each fold, threshold values were determined from the training data. Sensitivity, specificity, precision, classification error and Area Under Curve (AUC) from a Receiver Operator Curve (ROC) were calculated by applying the

thresholds to the testing data and were used as the performance metrics. Mean thresholds and performance metrics were calculated by averaging across the folds.

SDNR and RMSE threshold values were determined using sensitivity vs. specificity plots. For each fold, the SDNR threshold was varied over each SDNR value within the training dataset, and the sensitivity and specificity were calculated by applying the threshold to the training data. The optimal threshold was chosen as the value where sensitivity equalled specificity. This process was then repeated for the RMSE values.

Upper and lower thresholds were determined for FWHM and PSR. For each fold, the FWHM values of signal-time courses from the training data were ordered by increasing FWHM value. The signal-time courses with the smallest and largest FWHM values that passed qualitative review were identified and these values were used as the thresholds, with any signal-time course with an FWHM between the two thresholds classed as good quality. This process was repeated for the PSR values.

### 3.2.5. Combining Quantitative Measures Using Logical Voting

The threshold values from each of the quantitative measures were combined using a logical voting method and k-fold CV with k = 10. This process was carried out simultaneously alongside determining thresholds for each of the metrics. The thresholds determined from the training data during each fold were combined in a logical voting system and applied to the testing data. As with the individual thresholds, sensitivity, specificity, precision, classification error and AUC, were calculated and averaged across all folds.

In the logical voting system, the PSR and FWHM values from each signal-time course were compared to the thresholds. If a signal-time course failed one or both of the metrics it was classed as poor quality. If it passed both these metrics it was then compared to the SDNR and

RMSE thresholds. If the signal-time course passed at least one of these metrics it was classed as good quality. If it failed both metrics it was classed as poor quality.

### 3.2.6. Combining Quantitative Measures Using Machine Learning

Machine learning classification was carried out using the machine learning toolbox in Matlab (The MathWorks, MA, 2019a) and was applied to the quality control test dataset. Classifiers were trained with SDNR, RMSE, FWHM and PSR values used as predictors, and the qualitative review results used as the target outputs. Hyperparameter optimisation was applied to each classifier and k-fold CV with k = 10 was used. The classifiers used were binary tree, support vector machine (SVM), ensemble, random forest, and logistic regression. An explanation of how each machine learning classifier works can be found in Section 2.3.4.2, and a table summarising why each classifier is suited to the data is presented in Section 2.3.5. The binary tree classifier was selected as it is a simple and efficient classifier which has shown to be robust to outliers (295). The SVM classifier was selected as it has previously been shown to suit datasets where there are a limited number of features (266). The ensemble classifiers were used as they have been shown to offer improved performance compared to standard classifiers (271, 297). Five ensemble classifier methods were used, and these were AdaBoost, Gentle AdaBoost, Logit Boost, RUSBoost, and bagging. Hyperparameter optimisation was used to select the ensemble method used. As the weak learner was set to be a binary tree the bagging method produced a random forest. A separate random forest classifier was also trained and tested, as it allowed for more control over the hyperparameters. Gentle AdaBoost and Logit Boost were selected as they have been designed to give improved performance compared to AdaBoost and RUSBoost was selected as it is suited to datasets with class imbalances (297). Random forest was chosen as it is also a type of ensemble classifier and offers improved performance compared to a binary tree, it is also suited to smaller datasets (271, 297). The logistic regression classifier was chosen as it is a simple and

efficient classifier which is suited to binary classification problems (298). The average sensitivity, specificity, precision, classification error and AUC were calculated for each classifier.

### 3.2.7. Application to Patient Data

Each of the thresholds of the quantitative measures of quality and the best performing machine learning classifier were applied to one slice of patient data that was acquired using the acquisition protocol described in row 3 of Table 3.1 but was not included in the quality control test dataset. A quality map was created for each of the different methods, displaying which voxels had passed quality control and which had failed.

### 3.2.7. Statistical Analysis

For the subset of signal-time courses assessed by two reviewers, the agreement between reviewers was investigated by calculating the % disagreement between the two reviewers and the Cohen's Kappa for interrater reliability. All statistical analysis was carried out in R (R Foundation for Statistical Computing, Vienna, Austria, Version 3.5.0).

## 3.3. Results

### 3.3.1. Qualitative Review Results

Figure 3.1 shows an example signal-time course and identifies the features used to calculate each of the quantitative measures. Table 3.2 splits the signal-time courses assessed by two reviewers into 3 groups (all the signal-time courses, all the 1.5T signal-time courses, and all the 3T signal-time courses) and summarises the percentage disagreements and Cohen's Kappa for each group. Across the entire subset, the signal-time courses where there were disagreements had a median SDNR of 5.4 with a range of 3.3 - 56.4, median RMSE of 0.020 with a range of 0.005 - 0.058, median PSR of 84.5% with a range of 52.5% - 107.1%, and a median FWHM of 7s with a range of 4s - 15s.

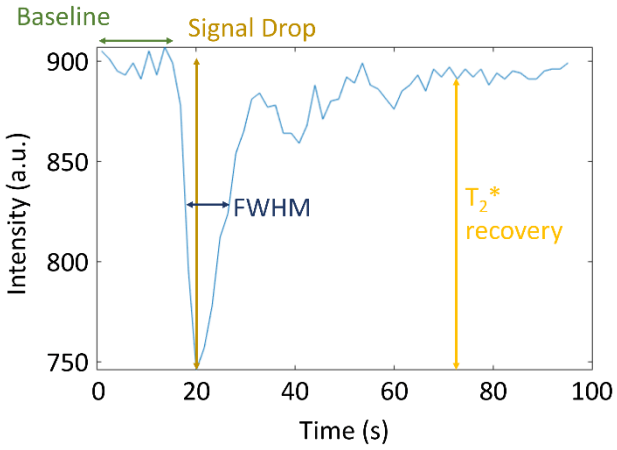



*Figure 3.1: Example signal-time course showing which features of the signal-time course are used to calculate the quantitative measures of quality.*

| Group | No. of Signal-Time Courses | Percentage Difference (%) | Cohen's Kappa |
|---|---|---|---|
| All | 243 | 6.58 | 0.84 |
| All 1.5T | 81 | 12.35 | 0.73 |
| All 3T | 162 | 3.70 | 0.79 |

*Table 3.2: A summary of the differences between the reviewers, in terms of percentage difference and Cohen's Kappa, across all signal-time courses, all signal-time courses recorded at 1.5T, and all signal-time courses recorded at 3T.*

### 3.3.2. Comparing Qualitative Review to Quality Metrics

Figure 3.2 shows some examples of signal-time courses that failed qualitative review for different reasons: 3(a) failed due to a small signal drop in comparison to the noise in the baseline, leading to a low SDNR; 3(b) failed due to a very noisy first pass, leading to a large RMSE value; 3(c) failed due to a very narrow first pass, which lead to a very small FWHM value; 3(d) failed to a low $T_2^*$ recovery, which lead to a very low PSR value. Below an SDNR value of 2.8 no signal-time courses passed qualitative review, whilst above an SDNR of 56.4, no signal-time courses failed qualitative review. Above an RMSE value of 0.0846 no signal-time courses passed qualitative review, whilst below an RMSE value of 0.0055 no signal-time courses failed qualitative review. No signal-time courses with FWHM values less than 3s or greater than 19s passed qualitative review. No signal-time courses with PSR values less than

43% or greater than 131% passed qualitative review. This shows that there is a region of uncertainty where it is difficult to classify signal-time courses.

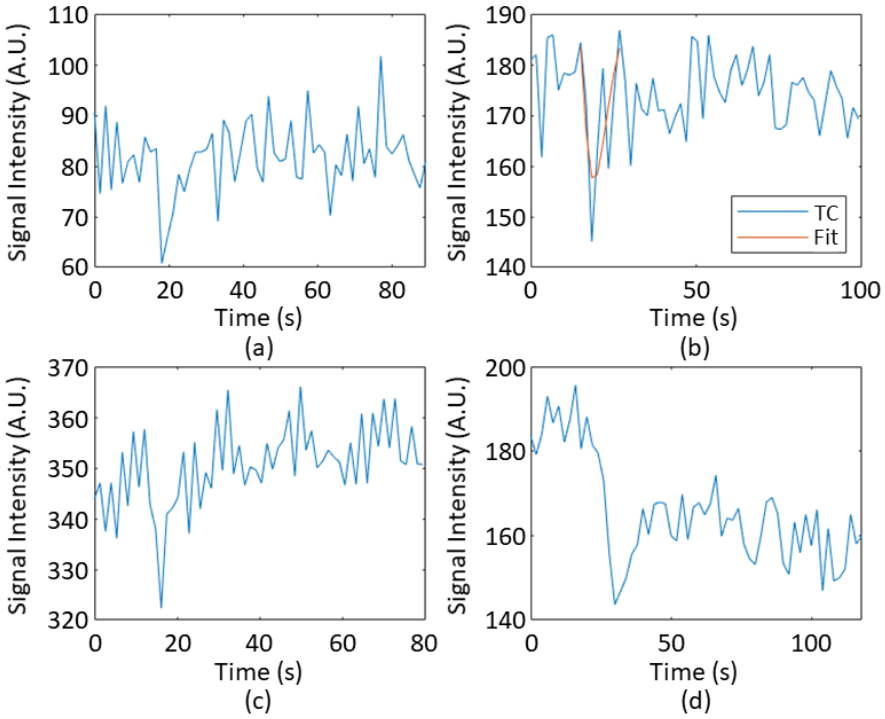



*Figure 3.2: An example of signal-time courses that failed qualitative review for differing reasons. (a) has a low SDNR value of 2.7, (b) has a large RMSE values of 0.129, (c) has a very small FWHM value of 1.8s and (d) has a low PSR of 36.9%.*

### 3.3.3. Determining Thresholds in Quantitative Measures of Quality using Qualitative Review

Figure 3.3 shows an example of the sensitivity vs. specificity plots for one of the folds. Averaging across all folds produced an SDNR threshold of 7.6, an RMSE threshold of 0.019, FWHM thresholds of 3s and 19s, and PSR thresholds of 42.9% and 130.4%. The average sensitivity, specificity, precision, classification error and AUC are summarised in Table 3.3. SDNR was the best performing quantitative measure producing a sensitivity, specificity, precision, classification error and AUC of 0.86, 0.86, 0.93, 14.2% and 0.83, respectively. Figure 3.4 shows some of the disagreements between the qualitative review results and the SDNR threshold. Out of the three signal-time courses that passed the SDNR threshold but failed qualitative review, all three passed the FWHM and PSR thresholds, and one passed the RMSE threshold. All three signal-time courses failed qualitative review because of issues

with the post-bolus signal, which was not picked up by SDNR. Out of the three signal-time courses that passed qualitative review but failed the SDNR threshold, one passed the RMSE threshold and two passed the FWHM and PSR thresholds. This shows that SDNR is key to data quality. However, relying on one metric to determine quality risks misclassification of signal-time courses.

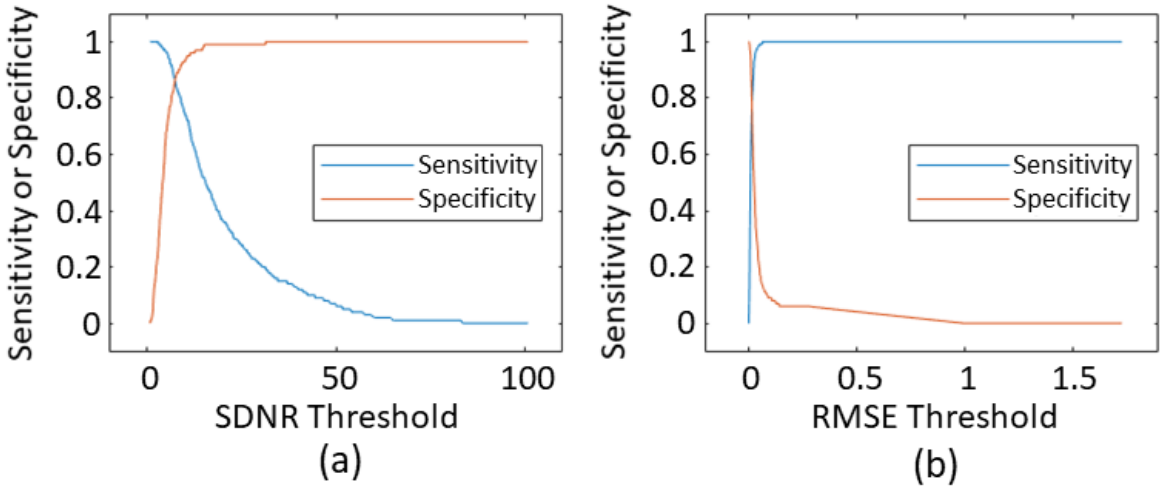



*Figure 3.3: Sensitivity and specificity plots for (a) SDNR, (b) RMSE.*

| Metric | Threshold(s) | Sensitivity | Specificity | Precision | Classification Error (%) | AUC |
|---|---|---|---|---|---|---|
| SDNR | 7.6 | 0.86 | 0.86 | 0.93 | 14.2 | 0.83 |
| RMSE | 0.019 | 0.79 | 0.79 | 0.89 | 21.8 | 0.75 |
| FWHM (s) | 3, 19 | 1.00 | 0.18 | 0.74 | 24.8 | 0.85 |
| PSR (%) | 42.9, 130.4 | 1.00 | 0.12 | 0.73 | 26.4 | 0.84 |

*Table 3.3: Summary of the threshold values, sensitivity, specificity, precision, classification errors and AUCs for each of the quantitative measures of quality.*

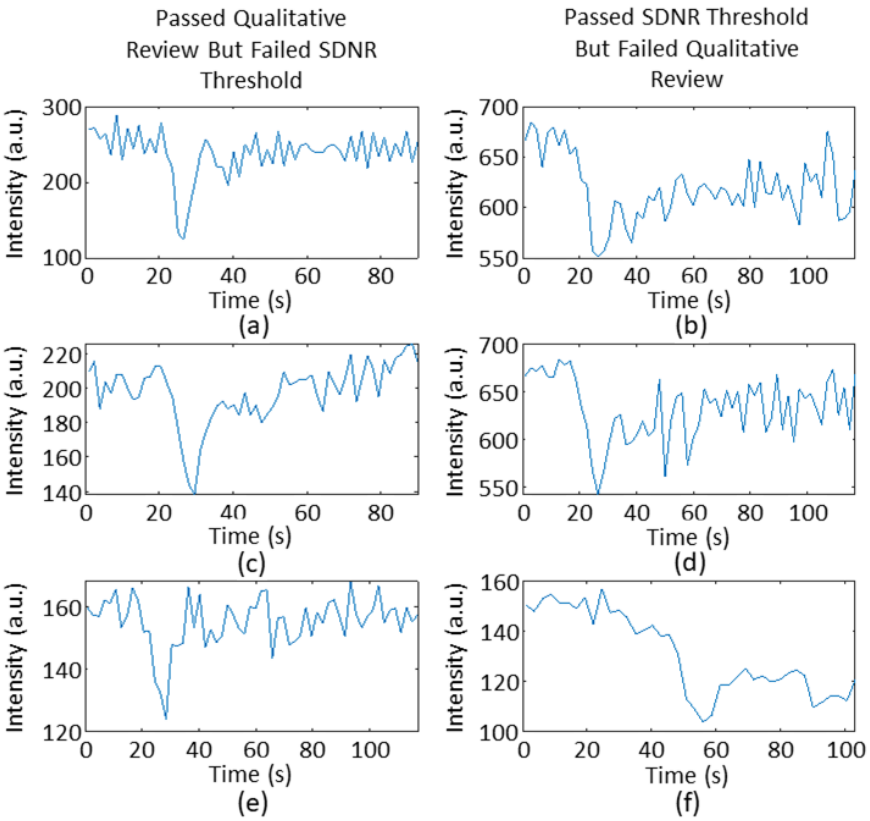



*Figure 3.4: A demonstration of the disagreements between the qualitative review results and the SDNR threshold. The left column (a, c, e) shows signal-time courses that passed qualitative review but failed the SDNR threshold. The signal-time courses in a, c, e has SDNR values of 5.4, 7.2 and 7.2, respectively. The right column (b, d, f) shows signal-time courses that passed the SDNR threshold but failed qualitative review. The signal-time courses in b, d, f had SNDR values of 7.8, 16.4 and 11.6, respectively.*

### 3.3.4. Combining Quantitative Measures Using Logical Voting

Figure 3.5 shows a summary of the logical voting process. Applying logical voting in combination with k-fold CV produced a sensitivity, specificity, precision, classification error and AUC of 0.91, 0.75, 0.89, 13.8%, and 0.84, respectively.

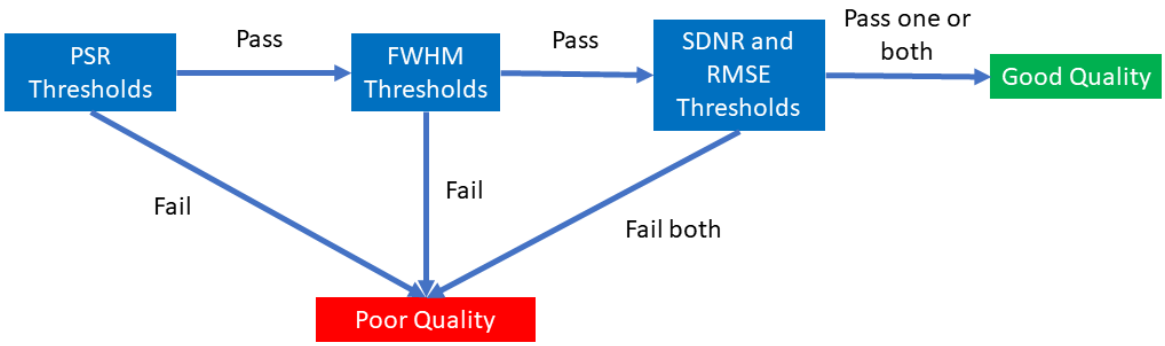



*Figure 3.5: Flowchart summarising the logical voting process.*

### 3.3.4. Combining Quantitative Measures Using Machine Learning

Table 3.4 summarises the sensitivity, specificity, precision, and classification error for each of the machine learning classifiers tested. The best performing classifier was the random forest, producing sensitivity, specificity, precision, classification error and AUC of 0.94, 0.83, 0.93, 9.3% and 0.89, respectively. Figure 3.6 shows examples of the disagreements between the qualitative review and machine learning results. Combining metrics to assess data quality offers improvements in classification error and machine learning classifiers offer an automated method to do this.

| Classifier | Sensitivity | Specificity | Precision | Classification Error (%) | AUC |
|---|---|---|---|---|---|
| Binary Tree | 0.93 | 0.80 | 0.92 | 11.4 | 0.87 |
| SVM | 0.93 | 0.84 | 0.93 | 9.5 | 0.89 |
| Ensemble* | 0.94 | 0.82 | 0.92 | 9.5 | 0.89 |
| Random Forest | 0.94 | 0.83 | 0.93 | 9.3 | 0.89 |
| Logistic Regression | 0.93 | 0.83 | 0.92 | 10.3 | 0.88 |

*Table 3.4: Summary of the performance metrics for each machine learning classifier. (* The Bag method was selected the by the hyperparameter optimisation).*

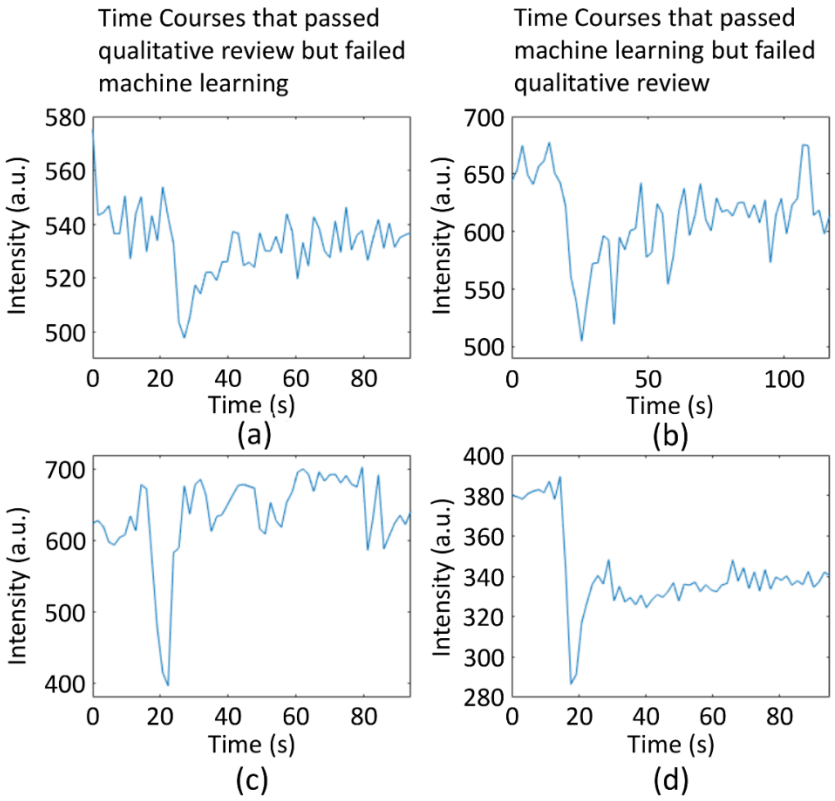


*Figure 3.6: An example of some of the agreements and disagreements between the machine learning results and qualitative review: (a & c) show signal-time courses that failed machine learning but passed qualitative review, (b & d) show signal-time courses that passed machine learning but failed qualitative review.*

### 3.3.5. Application to Patient Data

Figure 3.7 shows a set of quality maps produced by applying each of the quality control methods to one slice of a patient dataset from Birmingham Children's Hospital (acquired at 1.5T). In the quality maps blue pixels represent signal-time courses which passed the quality control methods, whilst orange pixels represent those that failed. Table 3.5 summarises the percentage of pixels passed by each method.

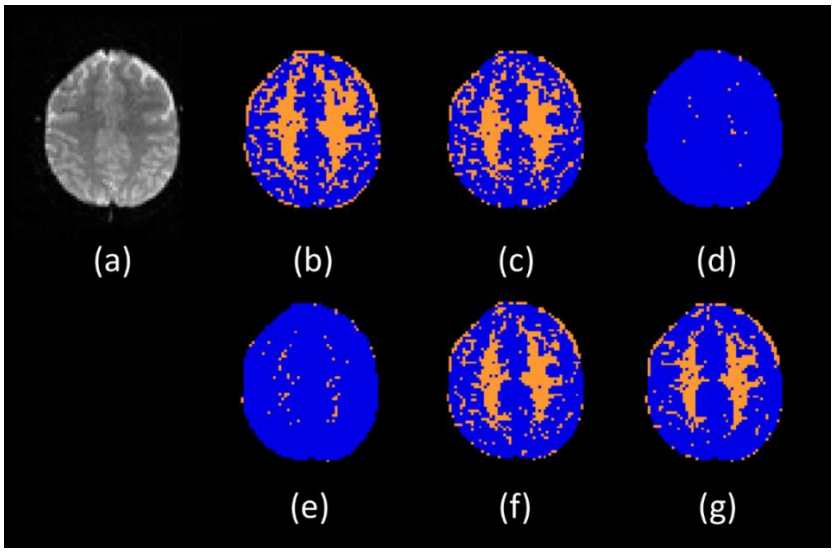



*Figure 3.7: (a) An axial slice from a patient dataset recorded at 1.5T, and the resulting quality maps from applying (b) the SDNR threshold, (c) RMSE threshold, (d) PSR threshold, (e) the FWHM threshold, (f) logical voting and (g) the machine learning classifier. Blue pixels represent signal-time courses that passed the quality control and orange pixels represent signal-time courses that failed quality control.*

| Quality Control Method | % of signal-time courses passed |
|---|---|
| SDNR Threshold | 60.97% |
| RMSE Threshold | 66.18% |
| PSR Threshold | 99.12% |
| FWHM Threshold | 97.44% |
| Logical Voting | 70.88% |
| Random Forest Machine Learning Classifier | 75.55% |

*Table 3.5: % of signal-time courses passed by each quality control method.*

## 3.4. Discussion

This study shows that although qualitative review can be used to assess data quality, there are a range of signal-time courses which are difficult to classify. Automated quality control methods can be developed using the results of qualitative review, and combining quantitative

measures using machine learning offers better performance than individual measures. However, selecting a set of quantitative measures to fully describe a signal-time course and all potential artefacts is challenging.

The signal-time courses assessed by two reviewers show a low discordance between reviewers, due to a low percentage of disagreements, and a Cohen's Kappa value of 0.83, which shows excellent agreement (300). Comparing this to the discordance between reviewers when the signal-time courses are split by field strength, it can be seen that the 1.5T signal-time courses are more difficult to classify. For the entire subset of signal-time courses where there were disagreements between the reviewers and for the signal-time courses that were within the region of uncertainty, the ranges of SDNR and RMSE values are both large. This shows that a signal-time course with a large SDNR is not guaranteed to be good quality, as other factors may affect quality. For example, a signal-time course with a low PSR, caused by $T_2^*$ leakage effects, could still have a large SDNR value.

The SDNR threshold defined the minimum acceptable SDNR and was the best performing individual measure, giving the most similar results to qualitative review. This is expected as SDNR defines how visible the signal drop is, which is a key part of assessing quality by qualitative review (34).

The RMSE threshold defined the maximum acceptable RMSE and gave poorer performance across all the performance measures compared to SDNR. This may be because RMSE focuses on the shape of the first pass, meaning signal-time courses that fail due to other artefacts could still have a reasonable RMSE. For example, RMSE is not able to detect anomalies in the post-bolus signal, which would be picked up by the PSR. A signal-time course which fails the PSR threshold could still have a good RMSE value.

The FWHM and PSR thresholds give AUC values comparable to the other metrics and a classification error similar to the RMSE threshold. They also give very good sensitivity and very poor specificity. However, this is due to the threshold selection method. Choosing thresholds based on the smallest and largest FWHM and PSR values that passed quality control means that there are very few false negatives and a lot of false positives. Unlike the SDNR and RMSE thresholds, FWHM and PSR define a range of acceptable values for each metric. FWHM has been previously shown to correlate to MTT, and as there is an expected range of values for MTT in GM and WM, there should be an expected range of values for FWHM (316). PSR is known to reflect the amount of contrast agent leakage, with low PSR values corresponding to $T_2^*$ weighted effects and a PSR values (greater than 100%) corresponding to $T_1$ weighted effects (31).

Multiple factors affect DSC-MRI data quality, and a single quantitative measure cannot cover all of them. For example, the right-hand column of Figure 3.4 shows signal-time courses that passed the SDNR threshold but failed qualitative review and at least one other threshold. The left-hand column of Figure 3.4 shows signal-time courses that failed the SDNR threshold but passed qualitative review and at least one other threshold. This, combined with the fact that it is difficult to define a single threshold for each quantitative measure, shows the need for combined quantitative measures to assess data quality. This agrees with work by Akella et al. where the failure rate of fitting a gamma-variate to the first pass, mean FWHM and mean PSR were calculated from each patient. Cut-off values were then calculated using a 99% one-sided confidence interval, and datasets that didn't fall within the cut-off values for at least one metric were classed as poor quality (307). This differs from the work presented here as thresholds for quality are determined using qualitative review results instead of confidence intervals.

Using logical voting to combine the thresholds offered an improved sensitivity, classification error and AUC compared to the SDNR threshold. However, this came at the cost of decreased specificity and precision. This shows that determining the optimal combination of thresholds manually is difficult.

Combining quantitative measures using machine learning classifiers offers an improved classification error compared to individual thresholds. They also provide a more efficient method for combining quantitative measures compared to logical voting, with improved performance. The random forest classifier gave the best classification error, and both the random forest and ensemble classifier gave better performance than the binary tree classifier, which is expected as these methods combine binary tree classifiers to give improved performance. For the ensemble method the hyperparameter optimisation selected the bag method as the best option. As the weak learner was a binary tree this was also a random forest. The separate random forest classifier gave slightly improved performance on the random forest produced by the ensemble method, and this may be because the separate random forest allowed for more control over the hyperparameter optimisation than the ensemble classifier. Of the three methods that did not combine classifiers (binary tree, SVM and logistic regression), the SVM had the best classification error. This is probably due to the fact that the SVM is suited to datasets with limited features and the dataset used in this work has four features (266). Although the random forest classifier gave the best classification error, it only offers a 0.2% decrease in classification error compared to the SVM and ensemble classifiers, whilst the rest of the performance measures give almost identical results. Therefore, any of these three classifiers would be suitable.

The machine learning classifier offers improved sensitivity, classification error and AUC, compared to the SDNR threshold. However, there is little change in specificity and precision.

This shows that the main improvement in performance comes from a reduction in the number of false negatives, with little change in the number of false positives. Machine learning also has a classification rate which is very similar to the percentage disagreements in qualitative review between reviewers, which implies that it is as accurate as qualitative review at least for these datasets. Therefore, when the quality control methods are applied to a patient dataset the machine learning classifier passes a higher % of signal-time courses than the SDNR threshold, as shown in Figure 3.6 and Table 3.5. The lack of reduction in the number of false positives is likely due to the current selection of quantitative measures not being able to identify all the artefacts that DSC-MRI is susceptible to.

When applied to patient data, the SDNR and RMSE thresholds and machine learning classifier have a similar effect to WM masking. GM has CBV values on average 2.7 times larger than WM (317), leading to a larger signal drop in the first pass of GM signal-time courses. The size of the signal drop also scales with field strength (34). Therefore, 1.5T WM data will have the lowest SDNR. As SDNR is key to data quality, the 1.5T WM data will generally be classified as poor quality compared to 3T data. However, there are multiple methods for measuring SDNR, therefore the method used may have an effect on the value of the SDNR.

The quality control methods presented in this work take a voxel-wise approach to assessing quality. Voxel-wise quality control would make it possible to exclude poor quality regions from a given dataset. However, this would be a problematic approach because, as discussed above, it could lead to the exclusion of large amounts of WM, which is suggested by consensus guidelines as the tissue to use for normalising rCBV values (318). This would affect the values of perfusion measures calculated from these datasets. It could also lead to the exclusion of useful information, such as pathology related signal-time courses, as the ML

classifier is trained using ‘normal brain’. A better approach would be to use voxel-wise quality control to give an overall assessment of data quality, which could then be used to decide whether a dataset should be included in a study. For example, if a dataset has a very low percentage of voxels passing quality control, then this could indicate the presence of significant artefact which would make the dataset unsuitable for use. However, if a dataset has a large percentage of voxels passing quality control then this dataset is likely to be suitable for inclusion. This allows for the quality of all the signal-time courses within a dataset to be assessed automatically, which is not currently possible using qualitative review. An alternative to this would be to assess the quality of an average signal-time course from a dataset. This would be much quicker than a voxel-wise analysis but could produce misleading results.

The majority of DSC-MRI studies use the ASFNR recommendation of qualitative review by eye to assess data quality (34), whilst some studies also use SNR (317). Automated quality control has previously been presented by Akella et al. (307) using statistical thresholds to assess quality. Our method differs as the thresholds and machine learning classifier are trained on the results of qualitative review. An alternative way to assess data quality is for a radiologist to assess the quality of the perfusion maps produced. This could either mean assessing the diagnostic quality of the perfusion maps (89, 319), or checking for the presence of susceptibility artefacts on the raw DSC images (320), or assessing the visibility of a certain region of the brain (321). However, these methods have not been automated and have the risk that artefacts may be misinterpreted as pathology. A difficulty of assessing the quality of the perfusion maps automatically is that the true perfusion values are not known. Therefore, the alternative is to assess the quality of the signal-time courses. As the perfusion parameters are

estimated from the signal-time courses, the quality of them will impact the quality of the perfusion parameters.

In order to calculate the measures presented in this study it is necessary to establish the end of the baseline. There are established methods for determining the end of the baseline, for example Carroll et al. present a method which uses adaptive thresholds calculated from the standard deviation of the pre-contrast signal (322). However, the method presented in this work is better suited to a multicentre dataset with variable injection protocols, as the Carroll method defines a set number of time points to calculate the adaptive thresholds from. This could be problematic in a multicentre study as the number of dynamics in the baseline varies between centres.

There are some limitations to this study. Firstly, the patient 'training' dataset does not include every type of artefact, such as motion or susceptibility artefacts. So, there may be cases where the classifier would misclassify a new signal-time course with an artefact not in the training set the first time. Secondly, these methods were tested on signal-time courses from slices of brain that did not contain tumour or other definite pathology. Quantitative measures may differ in diseased tissue and so the thresholds may not produce the same results in all circumstances. In particular, applying PSR to signal-time courses within brain tumours may not provide reliable results. Brain tumours can cause breakdown of the BBB, leading to contrast agent leakage. This causes local alterations in the $T_1$ and $T_2^*$, which affect the post bolus signal, altering the $T_2^*$ recovery of the signal-time courses (168). Leakage correction methods such as the Boxerman-Schmainda-Weisskoff method (156) are often applied. Finally, although the classifier offers an automated quality control method it is still trained on cases classified by qualitative review, which has an element of subjectivity to it. However,

any automated quality control method requires some form of human input and will therefore have some subjective judgement.

### 3.5. Conclusions

Qualitative review of individual signal-time courses by the two reviewers showed good agreement on the signal-time courses they assessed. Machine learning offers the possibility of a new automated alternative to qualitative review, which allows the quality of large datasets to be assessed, and combines multiple quantitative measures to give a reasonable classification error. Although SDNR is a good general indicator of quality, using only a single quantitative measure to determine data quality risks misclassification of signal-time courses with other types of artefact present. Combining SDNR with RMSE, FWHM and PSR improves classification and achieves a misclassification rate which is similar to the discordance rate of the qualitative review.

# Chapter 4: Denoising Methods to Improve Data Quality and Accuracy of Perfusion Measures for DSC-MRI

# 4. Denoising methods to improve data quality and accuracy of perfusion measures for DSC-MRI

## 4.1. Introduction

Estimating accurate measures of perfusion from DSC-MRI is challenging due to noise and other artefacts such as contrast agent leakage, motion, and susceptibility artefacts (34). Since perfusion measurements are estimated from the signal-time course, their accuracy is affected by the presence of noise and other artefacts. Acquisition parameters such as echo time (TE), repetition time (TR), flip angle, field strength, voxel volume and the protocol for the injection of the bolus of contrast agent can all affect the signal-to-noise ratio (SNR) (323).

Data quality of DSC-MRI data can be improved by application of post-processing methods, applied after the images have been acquired. An example of this is denoising, which could be used to suppress the noise in DSC-MRI data and therefore improve the accuracy of any perfusion measures (324, 325). There are many different methods available for denoising data, and this chapter focuses on matricization, wavelets and Tucker decomposition as potential denoising methods. Matricization has previously been used as an effective denoising method for MR images, for dynamic contrast enhanced (DCE-) MRI, and for MRI chemical shift imaging (CSI) (211, 214, 216, 326). Wavelet denoising has previously been applied to MRI of the liver, and to magnetic resonance spectroscopy (232, 240). Tucker decomposition has previously been applied to the denoising of hyperspectral images, and in magnetic resonance spectroscopy (MRS) (216, 227, 228).

The aim of this chapter is to validate the impact of several denoising methods on data quality using simulated DSC-MRI data, and then confirm this in patient data. In order to achieve this the objectives are to: apply matricization, wavelet denoising and the Tucker decomposition to simulated DSC-MRI signal-time courses; determine if there is an SDNR ‘performance

threshold' for each denoising method, below which CBV accuracy is not improved; compare the denoising methods to see which performs best in terms of noise suppression and CBV accuracy; and then to apply the denoising methods to patient data, to determine if the same level of noise reduction is achieved.

## 4.2. Methods

### 4.2.1 Simulations

DSC simulations were performed in Matlab (The MathWorks, MA, 2019a). First an arterial input function (AIF) was simulated using equation 4.1 (327). Following this a residue function was simulated using equation 4.2, and then the AIF and residue function were convolved to form a gadolinium concentration time curve (CTC) (328). There are several different models that can be selected for the residue function. An exponential model was chosen as it assumes that the vasculature is one, well-mixed compartment (328).

$$Cp(t) = \begin{cases} 0, & t < t_0 \\ K(t - t_0)^{\alpha} e^{-\frac{(t-t_0)}{\beta}}, & t \geq t_0 \end{cases} \quad (4.1)$$

$$y(t) = CBF \cdot e^{-\frac{t}{MTT}} \quad (4.2)$$

In equation 4.1 α, β and K are coefficients which affect the shape of the AIF, $C_p(t)$, and are assumed to have values of 3, 1.5 and 2 respectively (328). Whilst $t_0$ is the bolus arrival time and is assumed to have a value of 10 s (328). In equation 4.2 y(t) is the residue function. Average values for MTT in grey matter (GM) and white matter (WM) were calculated from the literature, which produced values of 5.11 s for GM and 5.28 s for WM (329-332). Individual CBV values were produced for each simulated signal-time course by obtaining values from a Gaussian distribution, using a mean of 5.10 ml/100ml and standard deviation of 0.5 ml/100ml for GM and a mean of 2.36 ml/100ml and standard deviation of 0.2 ml/100ml for WM. Average values were calculated from the values reported in the literature (329-333). CBF values were then calculated from the central volume theorem (equation 2.22) using MTT

as above. The CTC was converted into a signal-time course, according to the spoiled gradient echo signal equation, shown in equation 4.3 (334). This equation was chosen as gradient echo is the most frequently used sequence in DSC-MRI, and is the sequence used at our centre (323).

$$S(t) = M_0 \frac{\sin(\alpha)\left(1 - e^{-TR \cdot \mathrm{R1(C)}}\right)}{1 - \cos(\alpha)\, e^{-TR \cdot R1(C)}} e^{-TE \cdot R2^*(C)} \tag{4.3}$$

In equation 4.3, $M_0$ is the proton density of the tissue, α is the flip angle used in the DSC-MRI acquisition, TR is the repetition time, TE is the echo time and R1(C) and R2*(C) are the $T_1$ and $T_2^*$ relaxation rates which vary with concentration of contrast, C, and are given by equations 4.4 and 4.5, respectively. In these equations, R1(0) and R2*(0) are the $T_1$ and $T_2^*$ relaxation rates of the tissue without any contrast agent present, whilst r1 and r2 are the $T_1$ and $T_2^*$ relaxivities of the contrast agent, respectively (334).

$$R1(C) = R1(0) + r1 \cdot C \tag{4.4}$$

$$R2^*(C) = R2^*(0) + r2 \cdot C \tag{4.5}$$

R1(0) values for GM, WM and cerebrospinal fluid (CSF) were 0.92 $s^{-1}$, 1.29 $s^{-1}$ and 0.25 $s^{-1}$ respectively (335). R2*(0) values for GM, WM and CSF were 10.53 $s^{-1}$, 12.66 $s^{-1}$ and 0.4 $s^{-1}$ respectively (335). $M_0$ values were 79.8 pu, 69.2 pu and 100 pu for GM, WM and CSF (where pu is percentage units, the proton density as a percentage of the proton density of water) (336). The contrast agent relaxivities were set to the values of Magnevist contrast agent (GD-DPTA) in plasma at 37°C, which produced values of 4.9 $s^{-1}$ for r1 and 6.3 $s^{-1}$ for r2 (30). Values for TR, TE and α were selected using ASFNR recommendations and were assumed to be 1690 ms, 40 ms and 75˚ respectively (34).

Gaussian noise was added to signal-time course data, to give specified signal drop to noise ratio (SDNR) values. For each GM and WM signal-time course, the CBV was calculated

before noise was added, providing a ‘noiseless’ value for CBV, and allowing for the calculation of the % error in CBV of the noisy and denoised signal-time courses.

Two different models were created to test the denoising methods. The first of these was a simple model which consisted of a 100 x 100 grid of GM signal-time courses with differing amounts of Gaussian noise added to produce a model where each row had an average SDNR varying from 1 to 10. This model was used to test if it was possible to determine an SDNR performance threshold for the denoising methods.

The second model was a brain model that simulated a single axial slice of brain (matrix size 96 x 96), which contained a mixture of GM, WM, and CSF signal-time courses. This model was used to compare the performance of the different denoising methods. Three different version of this model were created by adding Gaussian noise to produce average SDNR values of 4, 8 and 12, respectively. For each signal-time course within each simulation, the SDNR, $\chi^2$ from comparison to the noiseless signal-time course and % error in CBV were calculated. % error in CBV was calculated from the % difference in CBV between the noiseless signal-time courses and the noisy and denoised signal-time courses. CBV values were calculated as the area under a gamma variate fitted to the concentration time curve. Figure 4.1(a) shows the simulated brain model and figure 4.2(b) shows the GM and WM signal-time courses used to construct the brain model.

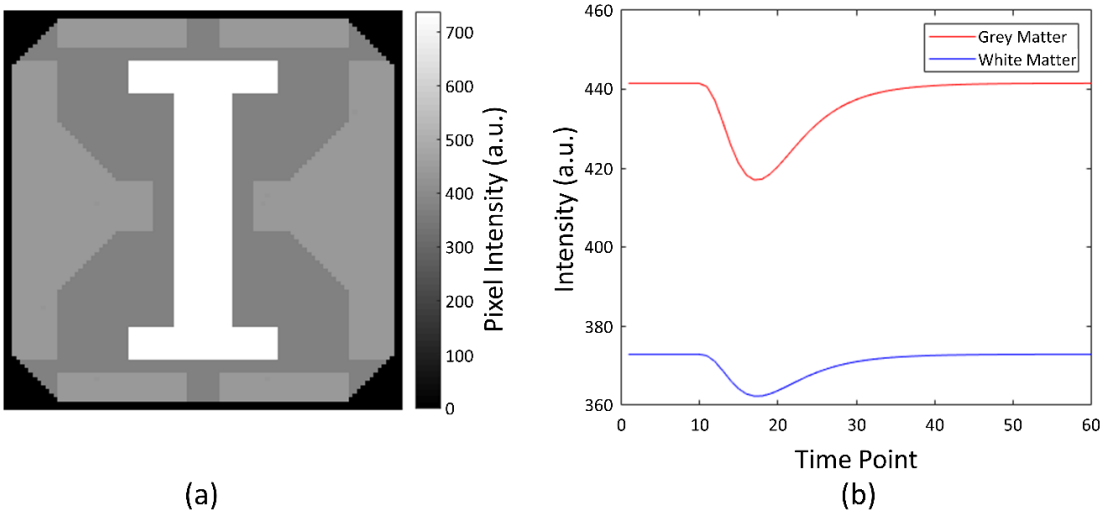



*Figure 4.1: (a) Simulated DSC-MRI dataset, containing a mixture of GM (light grey), WM (dark grey) and CSF (white). (b) Simulated Grey Matter (GM) and White Matter (WM) signal-time courses used in model. CSF signal-time courses are modelled as a flat horizontal line at an intensity of 739.*

### 4.2.2. Patient Data

Following testing of the denoising methods on simulated data, the methods were tested on a single slice from 8 different patient datasets. These patient datasets are from an imaging study entitled "CNS 2004 10 Functional Imaging of Tumours". Further details on this study can be found in section 3.2.1 of Chapter 3. Patients with cerebellar tumours were selected, and a slice was chosen from the corpus callosum to avoid any tissue containing tumour.

All 8 scans were acquired at Birmingham Children's Hospital. Four of the patient scans were acquired on a Siemens Avanto 1.5T (Siemens Healthcare, Erlangen, Germany) using the following parameters: TR = 1490 ms, TE = 40 ms, flip angle = 20°, voxel dimensions = 2.4 mm x 2.4 mm x 5 mm, scan type = GE-EPI, field of view = 230 mm x 230 mm, parallel imaging = SENSE with factor 2, temporal resolution = 1.49 s, number of dynamics = 60, total acquisition time = 90 s.

The other four patient scans were acquired on a Philips Achieva 3T TX (Philips Healthcare, Best, the Netherlands) using the following parameters: TR = 1865 ms, TE = 40 ms, flip angle = 20°, voxel dimensions = 2.5 mm x 2.5 mm x 3.5 mm, scan type = GE-EPI, field of view =

240 mm x 240 mm, parallel imaging = SENSE with factor 2, temporal resolution = 1.87 s, number of dynamics = 60, total acquisition time = 117 s.

All 8 patients were injected with Dotarem contrast agent (Dotarem, Guerbert, France), with a total dosage of 0.1 mmol/kg, using a power injector. All four patients scanned at 1.5T were injected using a single bolus protocol and received the total dosage in one injection at the start of time point 6 in the DSC data acquisition. All four patients scanned at 3T were injected using a split bolus protocol and received half the total dosage prior to the DSC acquisition, and the other half of the dosage at time point 6 in the DSC acquisition. This was to minimise $T_1$ weighted effects from contrast agent leakage (as discussed in section 2.2.2.2.). For all patients, a volume of up to 10 ml of saline was injected after the total dosage of contrast agent was received. All injections were administered at a rate of 3 ml $s^{-1}$.

### 4.2.3. Matricization Denoising

The matricization denoising method used a combination of matricization and PCA to denoise the data. An explanation of how both these techniques work can be found in section 2.3.1.1 and 2.3.1.2. Prior to denoising every signal-time course within was normalised using z-score normalisation (subtract the mean and divide by the standard deviation) (337). This method was chosen as it is the easiest to reverse, which is necessary when reconstructing data. Following normalisation, the data was matricized and decomposed using PCA. The number of components used to reconstruct a given dataset was selected by determining the elbow of the scree plot of the eigenvalues for that dataset (338). The location of the elbow was determined using the triangle thresholding method, whereby the elbow is defined by the maximum difference between the scree plot and a straight line between the first and last points of the scree plot (339). The data was then reconstructed, the normalisation reversed, and the denoised signal-time courses were folded back into a rank 3 tensor.

### 4.2.4. Wavelet Denoising

Wavelet denoising was carried out using the wdenoise function from the wavelet toolbox in Matlab (340). The wdenoise function suppresses noise using the method described in section 2.3.2.1 of the theory chapter, with the number of levels of denoising and the wavelet used defined by the user. The simple model was used to determine the optimal wavelet and level of denoising to use. The number of levels of denoising was fixed to one and the type of wavelet was varied through all those available from the wavelet toolbox and the wavelet that produced the lowest average % error in CBV was selected. This was determined to be the biorthogonal 3.9 wavelet. The simulation was then repeated with the type of wavelet fixed as the biorthogonal 3.9 wavelet and the levels of denoising varied between 1 and 5 (the maximum levels of denoising allowed by wdenoise). From this simulation the level of denoising with the lowest % error in CBV was 3. Therefore, the biorthogonal 3.9 wavelet with 3 levels of denoising was used.

### 4.2.5. Tucker Decomposition Denoising

The process behind Tucker decomposition is explained in section 2.3.1.3. Tucker decomposition denoising can be applied to DSC-MRI by considering one slice of data as a rank 3 tensor, with the first 2 dimensions being the rows and the columns of the image and the third dimension being time. Denoising was carried out using the N-way and tensorlab toolboxes in Matlab (221, 341). The mlrankest function from the tensorlab toolbox was used to estimate the rank of the core tensor and this was used as an estimate for the number of components used in Tucker decomposition.

### 4.2.6. Determining a Performance Threshold in SDNR for Denoising Using the Simple Model

SDNR and CBV error values were investigated to see if there was a performance threshold for denoising, a minimum pre-denoising SDNR value below which there was no improvement in CBV error post-denoising. All three denoising methods were independently applied to the

simple model and the CBV error post-denoising was compared to the CBV error pre-denoising. The signal-time courses were then split into two groups: those with improved CBV error post-denoising, and those with worse CBV error post-denoising. The SDNR values pre-denoising of each group were plotted as histograms to investigate the link between pre-denoising SDNR and change in CBV error.

### 4.2.7. Comparing Denoising Methods using the Brain Model

Denoising methods were compared using the three different versions of the brain model with average SDNR values of 4, 8, and 12, respectively. Prior to denoising, SDNR, $\chi^2$, and % error in CBV values were calculated for each of the brain models. Each of the three denoising methods were then applied independently to each of the brain models, and the SDNR, $\chi^2$ and % error in CBV values were recalculated post-denoising. Violin plots were produced to graphically demonstrate the differences in distributions of SDNR, % error in CBV and $\chi^2$ values. CBV maps were produced pre- and post-denoising to show the effect of denoising on the perfusion parameters.

### 4.2.8. Comparing Denoising Methods in Patient Data

SDNR values were calculated for the signal-time courses within the selected slice from each patient dataset. Each of the denoising methods were then applied individually to each slice of patient data and the SDNR values were recalculated. Violin plots were used to demonstrate the differences in SDNR values due to each denoising method. rCBV maps were produced pre-and post-denoising to show the effect of denoising on the perfusion parameters.

### 4.2.9. Statistics

Statistical analysis was carried out using R (R Foundation for Statistical Computing, Vienna, Austria, Version 3.5.0), and was used to compare differences in SDNR, % error in CBV and $\chi^2$ values pre- and post-denoising in the brain model; and differences in SDNR pre- and post-denoising in patient data. The Shapiro-Wilk test was used to test for the normality of the data,

and the Kruskal-Wallis and Wilcoxon pairwise tests were carried out to assess if there were any statistically significant differences between the metrics produced by each denoising method. A significance level of $P < 0.05$ was used, with a Bonferroni correction applied to the P values from the Wilcoxon pairwise test to account for the number of comparisons.

## 4.3. Results

### 4.3.1. Determining a Performance Threshold in SDNR Denoising using the Simple Model

Histograms of the pre-denoising SDNR values from the simple model were split by whether the post-denoising CBV error improved or worsened for each of the denoising methods. The histograms for matricization, wavelet denoising and Tucker decomposition are shown in Figures 4.2 (a), (b) and (c) respectively. On average there was an improvement in CBV error post denoising for each denoising method. There was an improvement in CBV error post-denoising in 73.9% of signal-time courses for matricization, 67.1% of signal-time courses for wavelet denoising, and 66.8% of signal-time courses for Tucker decomposition. Table 4.1 summarises the average change in % CBV error ± the standard deviation for each of the histograms in Figure 4.2. As there was an average improvement in CBV error across all SDNR values, it was not possible to determine a performance threshold for SDNR.

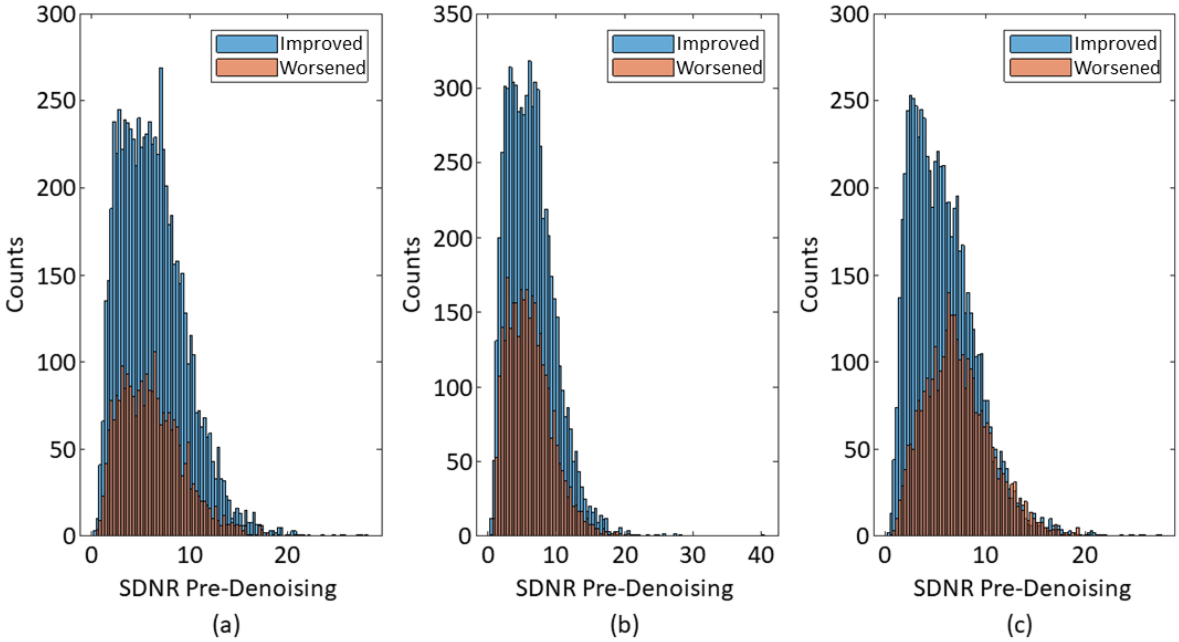



*Figure 4.2: Histograms of SDNR values from simulated signal-time courses split by whether they had improved CBV error (blue) or worse CBV error (orange) following (a) matricization, (b) wavelet denoising and (c) Tucker decomposition.*

| Denoising Method | Average decrease in CBV error post-denoising for signal-time courses with decreased CBV error | Average increase in CBV error post-denoising for signal-time courses with increased CBV error | Average change in CBV error post-denoising across all signal-time courses (negative value denotes average decrease) |
|---|---|---|---|
| Matricization | 13.1% ± 16.8% | 6.5% ± 7.4% | -8.0 ± 17.2 |
| Wavelet Denoising | 2.8% ± 4.4% | 3.9% ± 5.1% | -0.2 ± 5.6 |
| Tucker Decomposition | 16.3% ± 21.4% | 7.3% ± 5.9% | -8.5 ± 21.0 |

*Table 4.1: A table summarising the average change in CBV error for each of the signal-time courses within each of the histograms shown in Figure 4.2.*

### 4.3.2. Comparing Denoising Methods using the Brain Model

Figures 4.3 – 4.5 show example GM and WM signal-time courses from the brain models with SDNR values of 4, 8, and 12, respectively. Signal-time courses are shown before and after noise was added and after each of the denoising methods has been applied. Table 4.2 shows the number of components selected by the matricization and Tucker decomposition methods for each brain model.

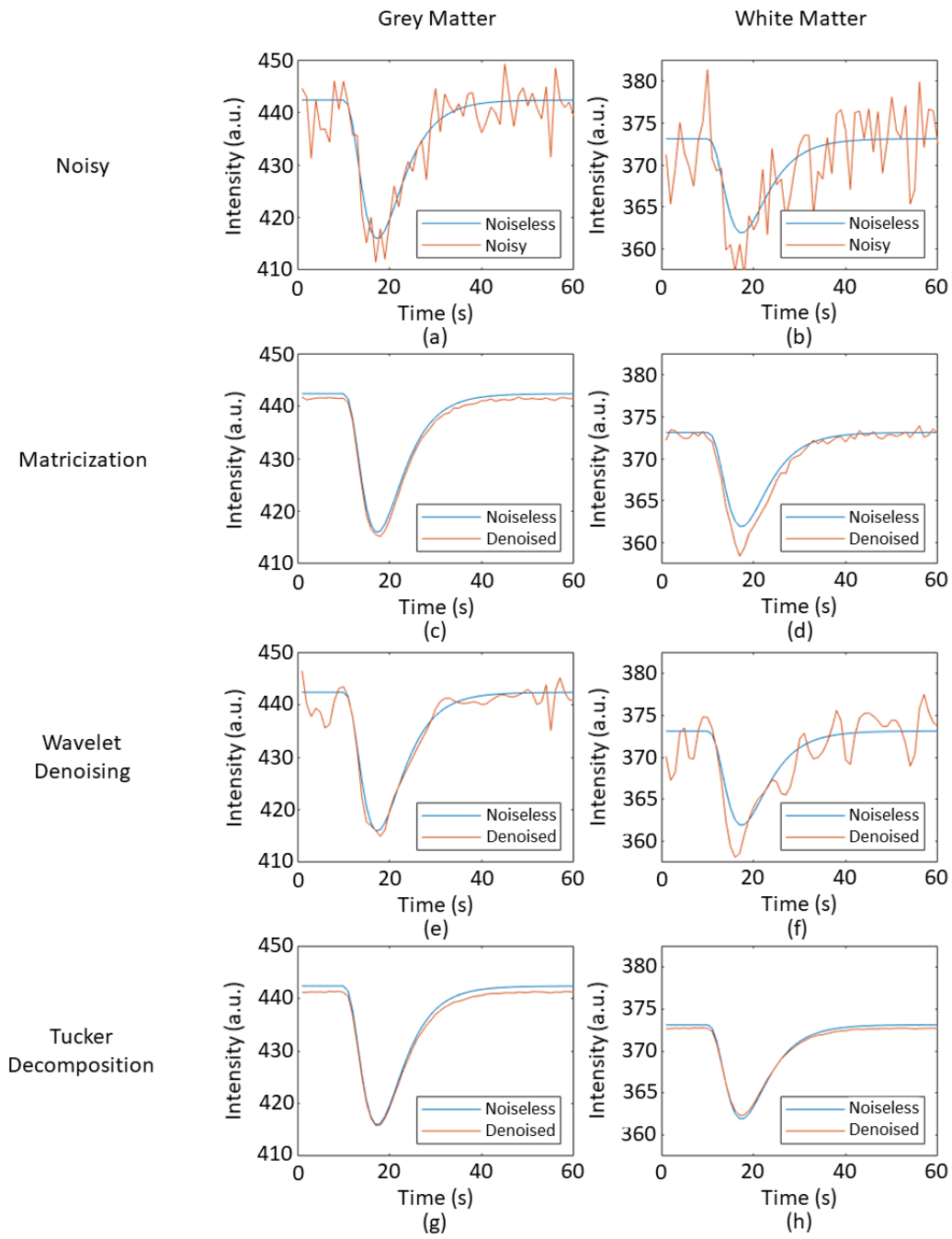



*Figure 4.3: Example of grey matter (left column) and white matter (right column) signal-time courses before and after denoising methods have been applied for the brain model with an average SDNR of 4. (a) GM signal-time course prior to denoising, (b) WM signal-time course prior to denoising, (c) GM signal-time course after matricization, (d) WM signal-time course after matricization, (e) GM signal-time course after wavelet denoising, (f) WM signal-time course after wavelet denoising, (g) GM signal-time course after Tucker decomposition, (h) WM signal-time course after Tucker decomposition.*

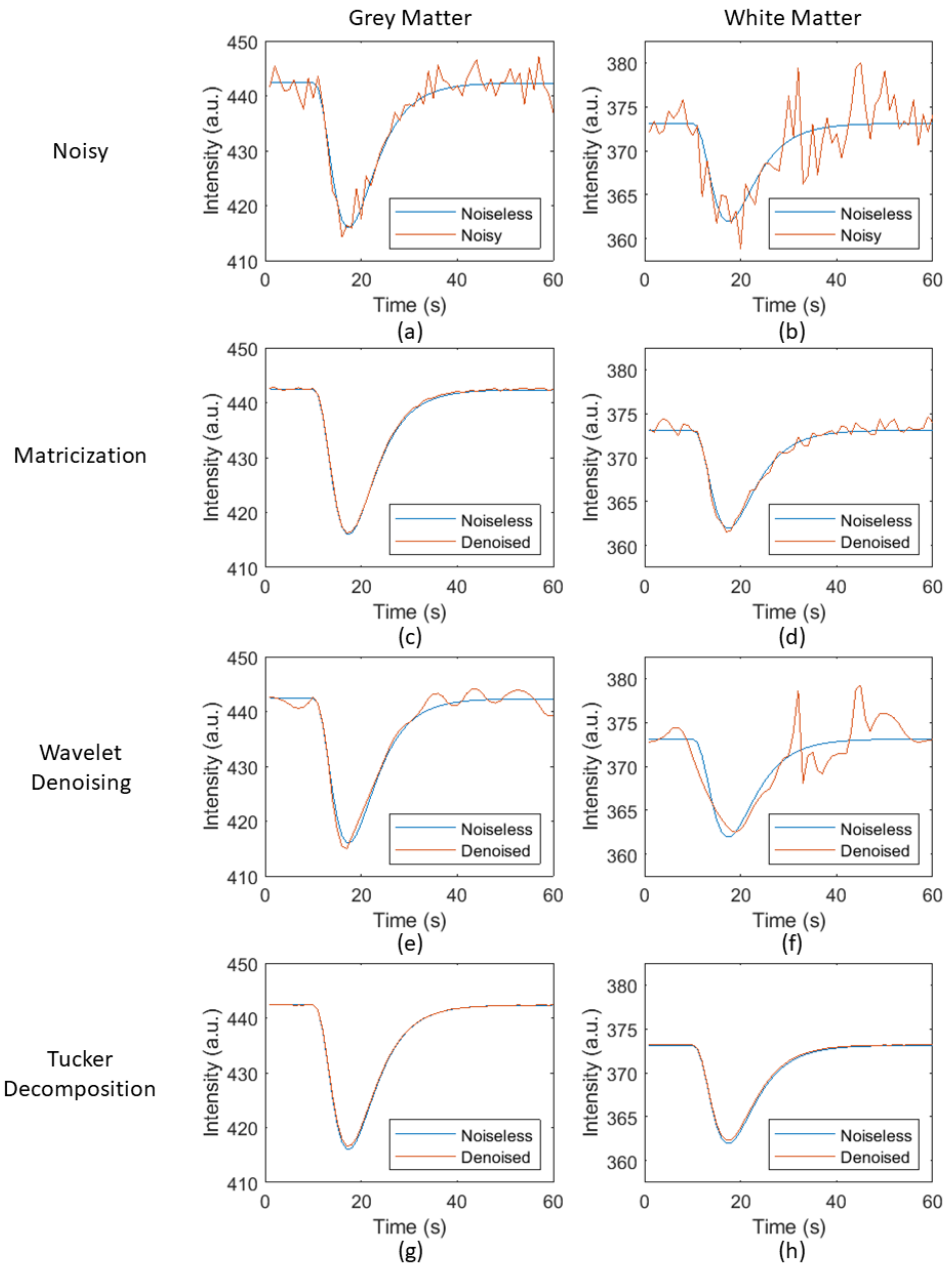



*Figure 4.4: Example of grey matter (left column) and white matter (right column) signal-time courses before and after denoising methods have been applied for the brain model with an average SDNR of 8. (a) GM signal-time course prior to denoising, (b) WM signal-time course prior to denoising, (c) GM signal-time course after matricization, (d) WM signal-time course after matricization, (e) GM signal-time course after wavelet denoising, (f) WM signal-time course after wavelet denoising, (g) GM signal-time course after Tucker decomposition, (h) WM signal-time course after Tucker decomposition.*

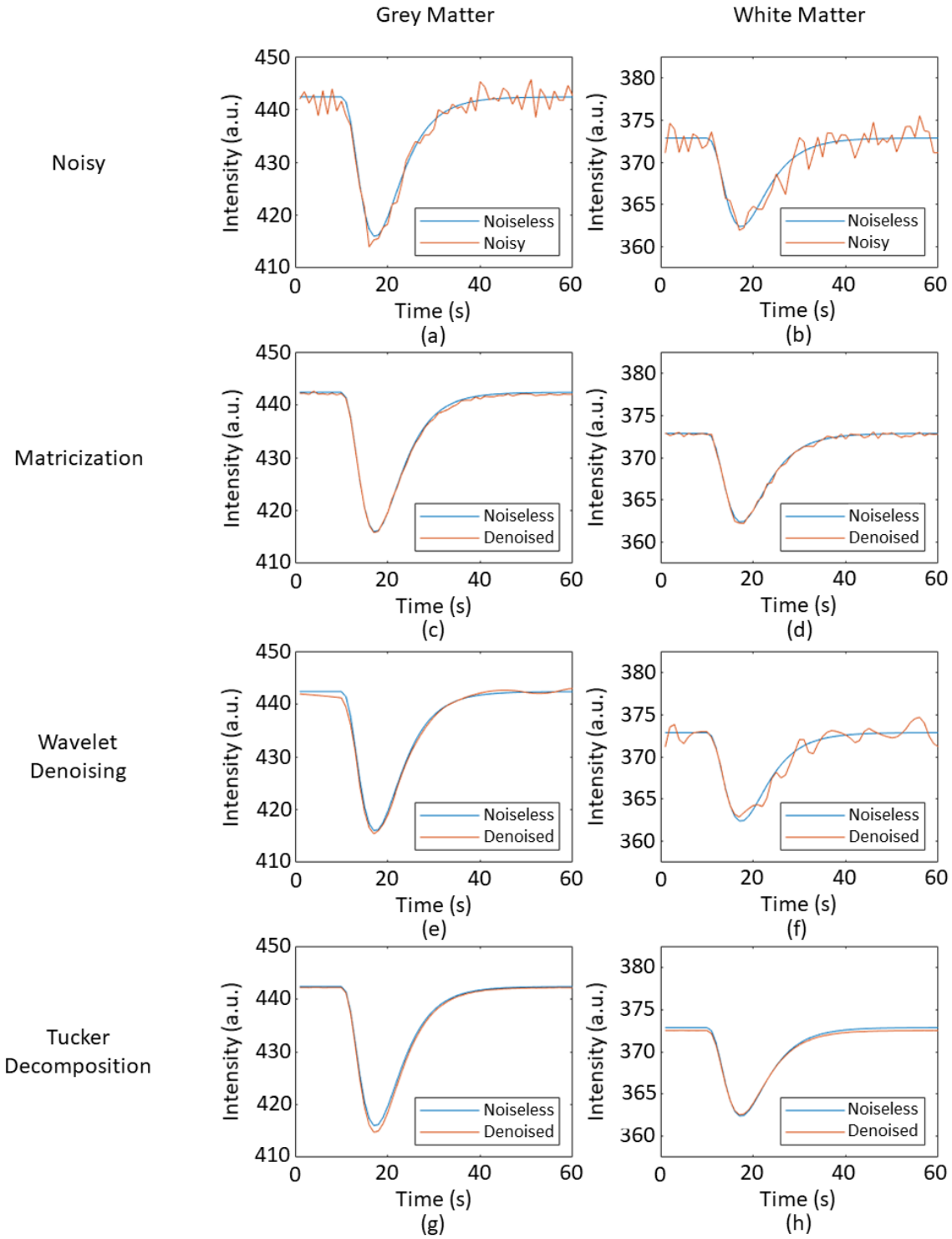



*Figure 4.5: Example of grey matter (left column) and white matter (right column) signal-time courses before and after denoising methods have been applied for the brain model with an average SDNR of 12. (a) GM signal-time course prior to denoising, (b) WM signal-time course prior to denoising, (c) GM signal-time course after matricization, (d) WM signal-time course after matricization, (e) GM signal-time course after wavelet denoising, (f) WM signal-time course after wavelet denoising, (g) GM signal-time course after Tucker decomposition, (h) WM signal-time course after Tucker decomposition.*

| Model | No. of Components | |
|---|---|---|
| | Matricization | Tucker Decomposition |
| SNR 4 | 2 | [32,32,2] |
| SNR 8 | 3 | [32,32,2] |
| SNR 12 | 3 | [32,32,2] |

*Table 4.2: The number of components selected by the matricization and Tucker decomposition denoising methods for each model.*

Figures 4.6 - 4.8 shows the distributions of each metric, pre- and post-denoising for all simulations. The Shapiro-Wilk test for normality gave $P < 0.05$ for each of the metrics shown in Figures 4.6 - 4.8, showing that all metrics are not normally distributed.

Figure 4.6 shows violin plots of the SDNR values. For all three brain models, matricization, wavelet denoising and Tucker decomposition all give significantly improved SDNR with $P < 0.05$. Tucker decomposition is significantly better than matricization and wavelet denoising, with $P < 0.05$, producing SDNR values of 285 ± 15, 575 ± 36, and 1223 ± 22 for the brain models with average SDNR values of 4, 8 and 12, respectively.

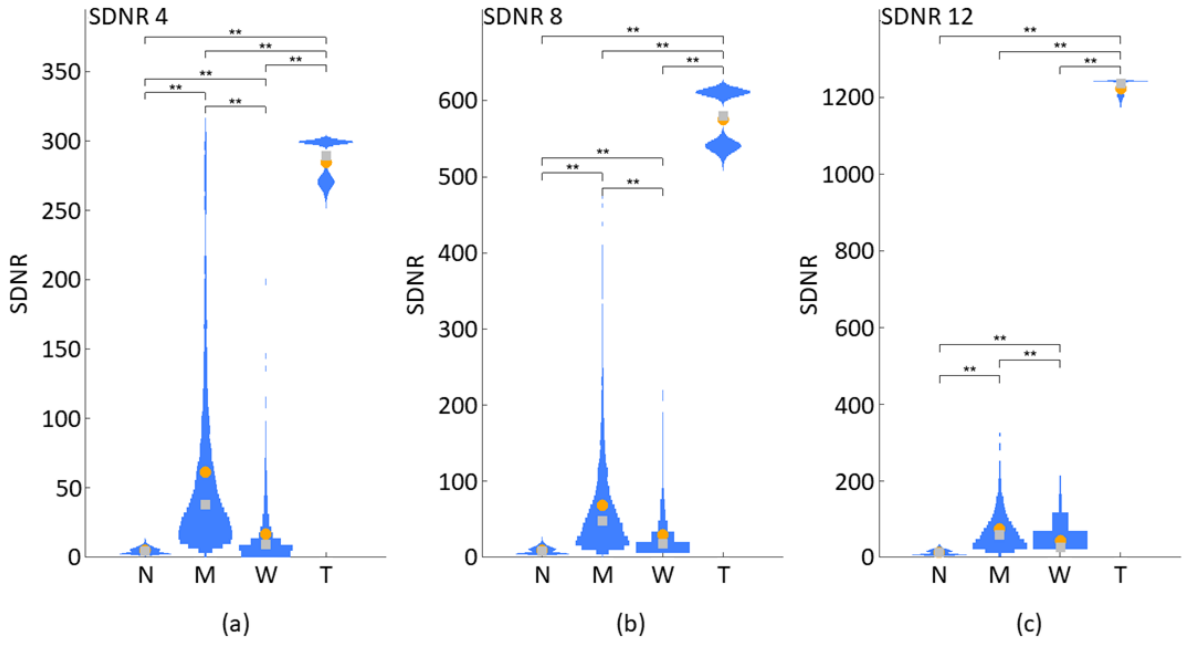



*Figure 4.6: The violin plots of SDNR for noisy data (N), data following matricization (M), data following wavelet denoising (W) and data following Tucker decomposition (T). (a) shows the brain model with an average SDNR of 4, (b) shows the brain model with an average SDNR of 8, and (c) shows the brain model with an average SDNR of 12. Brackets with * denote $P < 0.05$, and brackets with ** denote $P < 0.001$.*

Figure 4.7 shows violin plots of the $\chi^2$ values. For all three brain models, matricization, wavelet denoising and Tucker decomposition all give significantly improved $\chi^2$ with $P < 0.05$. For the brain model with an average SDNR of 4, Tucker decomposition offers significantly better $\chi^2$ than matricization and wavelet denoising ($P < 0.05$), with a value of $0.992 \pm 0.016$. Whilst for the brain models with an average SDNR of 8 and 12, matricization offers significantly better $\chi^2$ than wavelet denoising and Tucker decomposition ($P < 0.05$), producing values of $0.996 \pm 0.004$ and $0.998 \pm 0.002$, respectively.

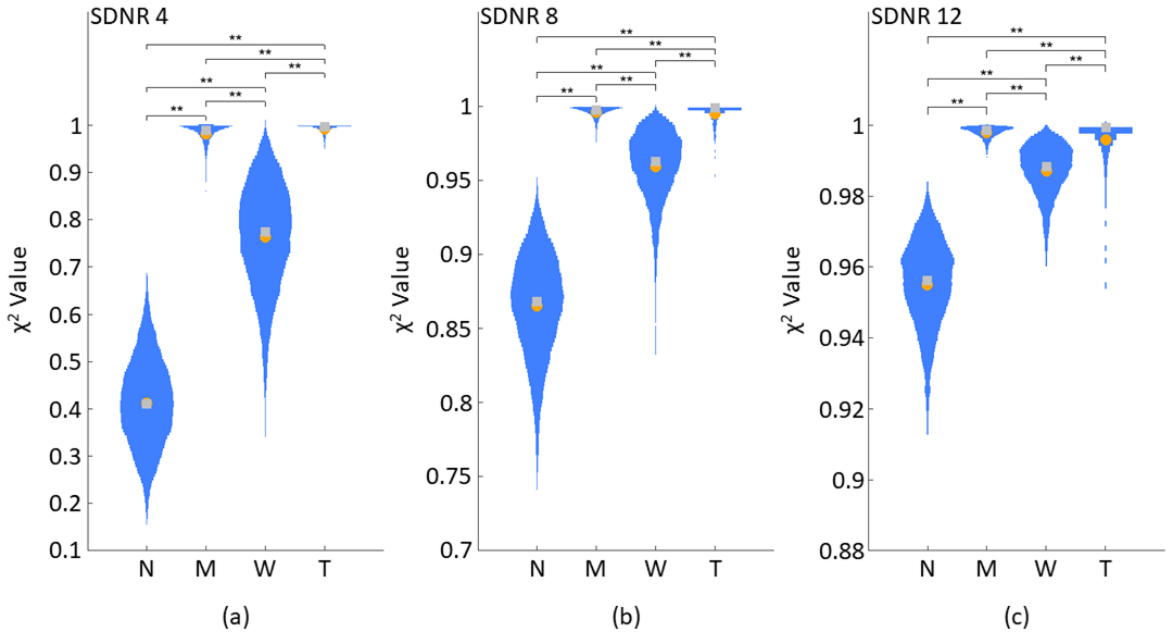



*Figure 4.7: The violin plots of $\chi^2$ for noisy data (N), data following matricization (M), data following wavelet denoising (W) and data following Tucker decomposition (T). (a) shows the brain model with an average SDNR of 4, (b) shows the brain model with an average SDNR of 8, and (c) shows the brain model with an average SDNR of 12. Brackets with * denote $P < 0.05$, and brackets with ** denote $P < 0.001$.*

Figure 4.8 shows the violin plots for the CBV error values. For the brain models with average SDNR values of 4 and 8, both matricization and Tucker decomposition give significantly better CBV error than the noisy data, with $P < 0.05$. However, for the brain model with an average SDNR of 12 matricization gives better CBV error than noisy data ($P < 0.05$), whilst Tucker decomposition gives worse CBV error than noisy data ($P < 0.05$). For the brain model with an average SDNR of 4, Tucker decomposition offers significantly better CBV error than

matricization ($P < 0.05$) with a value of 7.8% ± 6.2%. Whilst for the brain model with an average SDNR of 8, matricization offers significantly better CBV error than Tucker decomposition ($P < 0.05$), producing values of 5.4% ± 5.1% and 5.1% ± 4.0% respectively. Wavelet denoising does not offer a significant improvement in CBV error across all three simulations ($P = 1$).

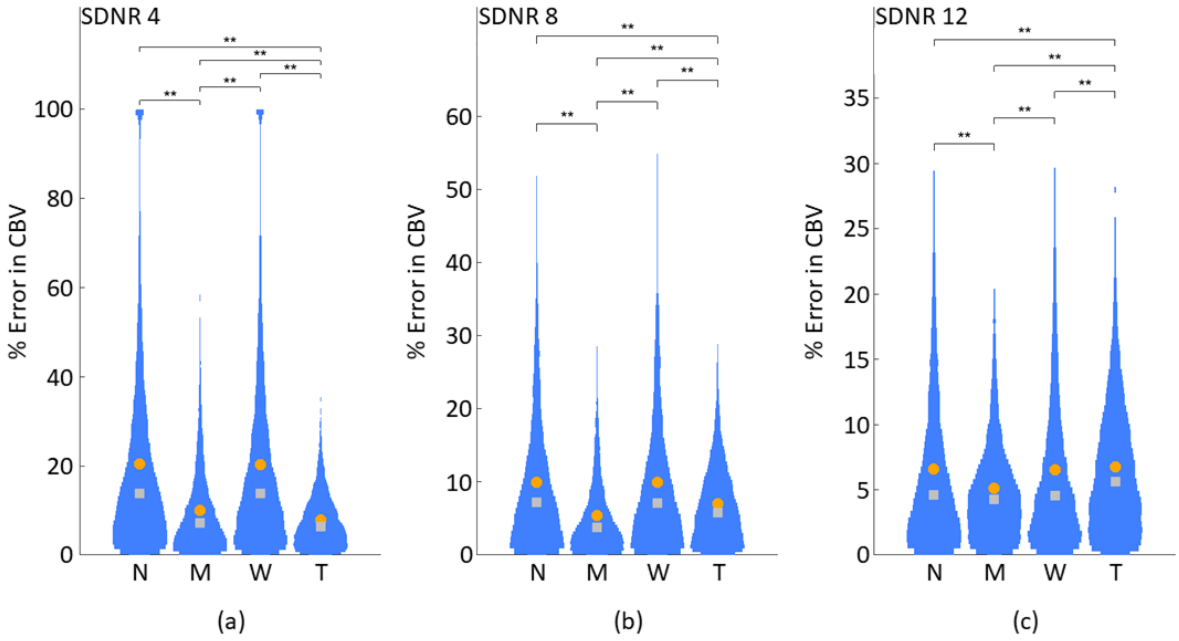



*Figure 4.8: The violin plots of % error in CBV for noisy data (N), data following matricization (M), data following wavelet denoising (W) and data following Tucker decomposition (T). (a) shows the brain model with average SDNR of 4, (b) shows the brain model with average SDNR of 8, and (c) shows the brain model with an average SDNR of 12. Brackets with * denote $P < 0.05$, and brackets with ** denote $P < 0.001$.*

Table 4.3 summarises the mean values ± the standard deviations for each metric, denoising method and simulation.

| | SDNR | | | $\chi^2$ | | | CBV Error (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sim 4 | Sim 8 | Sim 12 | Sim 4 | Sim 8 | Sim 12 | Sim 4 | Sim 8 | Sim 12 |
| Noisy | 4.8 ± 2.4 | 9.0 ± 4.8 | 13.3 ± 7.0 | 0.413 ± 0.092 | 0.865 ± 0.035 | 0.955 ± 0.012 | 20 ± 21 | 10.0 ± 9.3 | 6.6 ± 6.3 |
| Matricization | 61 ± 63 | 69 ± 66 | 75 ± 62 | 0.982 ± 0.022 | 0.996 ± 0.004 | 0.998 ± 0.002 | 10.1 ± 9.6 | 5.4 ± 5.1 | 5.1 ± 4.0 |
| Wavelet Denoising | 17 ± 22 | 29 ± 44 | 43 ± 93 | 0.763 ± 0.111 | 0.959 ± 0.023 | 0.987 ± 0.008 | 20 ± 21 | 9.9 ± 9.3 | 6.5 ± 6.2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Tucker Decomposition | 285 ± 15 | 575 ± 36 | 1223 ± 22 | 0.992 ± 0.016 | 0.995 ± 0.011 | 0.996 ± 0.011 | 7.8 ± 6.2 | 7.0 ± 5.5 | 6.8 ± 5.3 |

*Table 4.3: Table showing the mean ± standard deviation in SDNR, $\chi^2$ and % error in CBV values before and after denoising for each of the brain models.*

Figure 4.9 shows the rCBV maps for each of the simulated models prior to denoising and after each of the denoising methods had been applied.

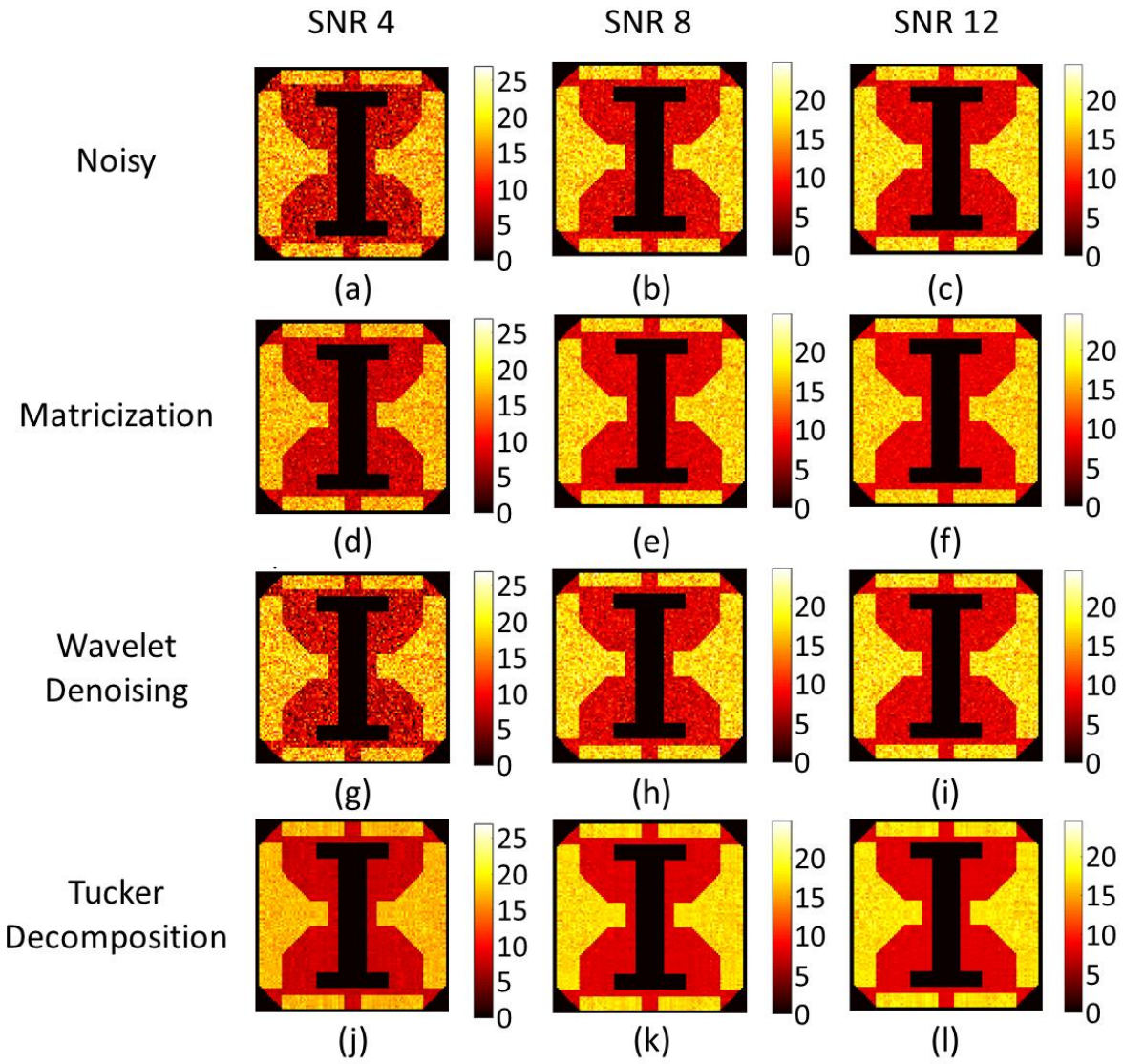



*Figure 4.9: The rCBV maps produced for the simulated data pre- and post-denoising. (a-c) shows the rCBV maps pre-denoising for each of the three models; (d-f) shows the rCBV maps post matricization denoising for each of the three models; (g-i) shows the rCBV maps post wavelet denoising for each of the three models; and (j-l) shows the rCBV maps post Tucker decomposition denoising for each of the three models.*

Figure 4.10(a) shows a simulated signal-time course prior to denoising, which has a CBV error 3.80%. Figure 4.10(b) shows the signal-time course after matricization, which has an improved CBV error of 3.52%. Figure 4.10(c) shows the signal-time course after wavelet denoising, which has a CBV error of 5.62%. Figure 4.10(d) shows the signal-time course after Tucker decomposition, which has a CBV error of 7.22%.

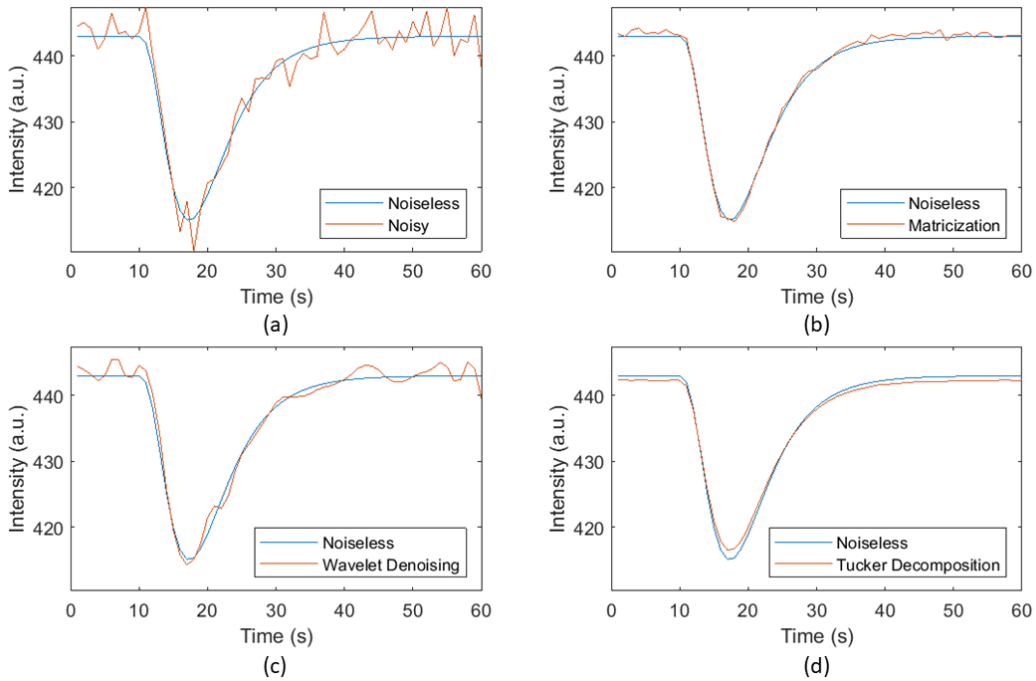



*Figure 4.10: An example of denoising a simulated signal-time course with initial SDNR of 19.9. In this example Tucker decomposition removes some of the first pass as well as denoising. (a) shows the signal-time course pre-denoising, (b) shows the signal-time course after matricization, (c) shows the signal-time course after wavelet denoising and (d) shows the signal-time course after Tucker decomposition.*

### 4.3.3. Comparing Denoising Methods on Patient Data

Figure 4.11 shows an example of each of the denoising methods applied to a signal-time course from a 3T and a 1.5T patient dataset. The signal-time course from the 3T patient dataset had an original SDNR of 21.2 and SDNR values of 64.3, 17.1 and 58.8 after matricization, wavelet denoising and Tucker decomposition, respectively. The signal-time course from the 1.5T patient dataset had an original SDNR of 24.9 and SDNR values of 128.4, 42.4 and 189.1 after matricization, wavelet denoising and Tucker decomposition, respectively.

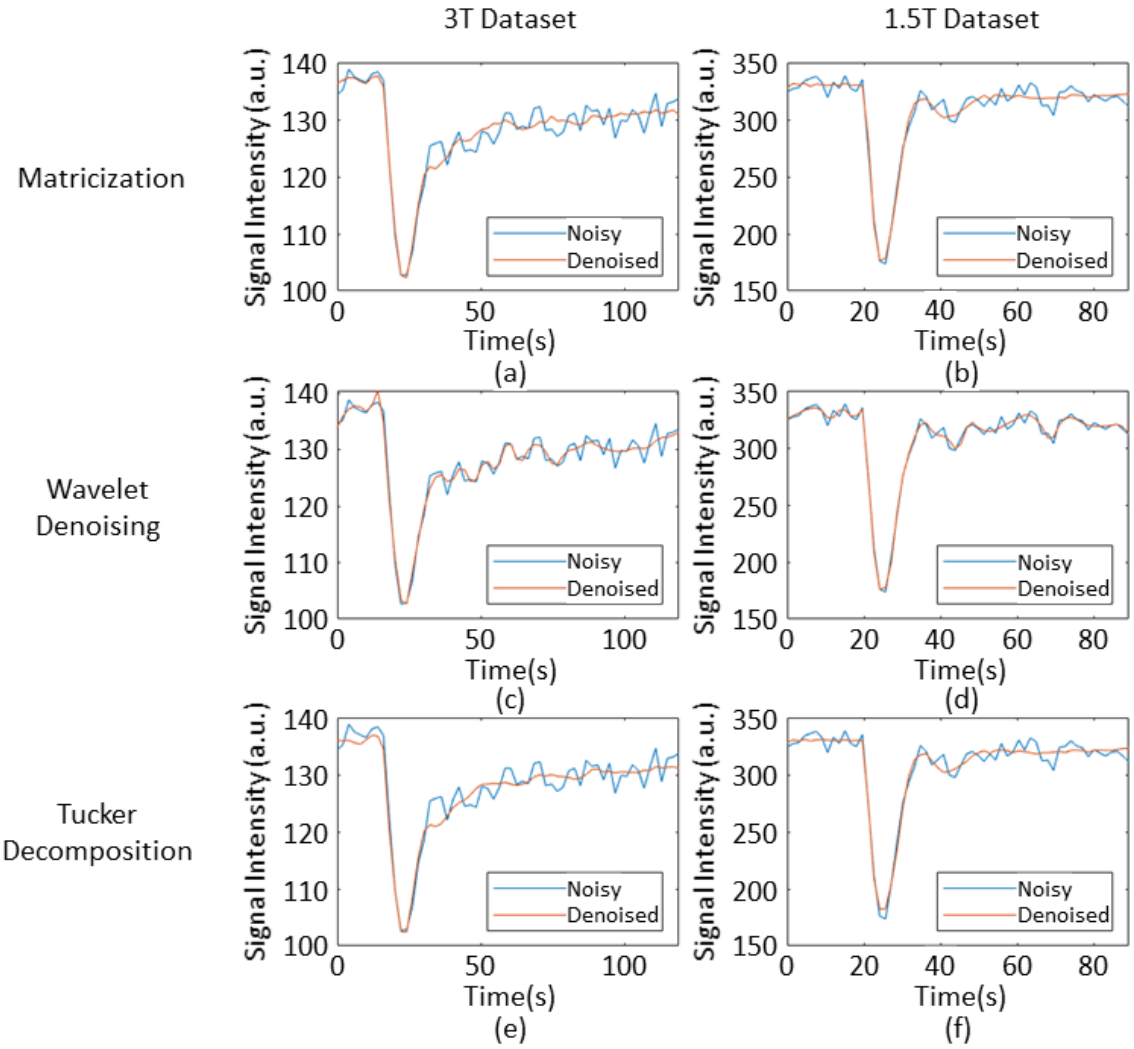



*Figure 4.11: An example of a signal-time course from a 3T patient dataset (left column), with original SDNR of 21.2, and an example of a signal-time course from some 1.5T patient dataset (right column), with original SDNR 24.9, after matricization (top row), wavelet denoising (middle row) and Tucker decomposition (bottom row) has been applied. (a) Matricization applied to a signal-time course from the 3T dataset, giving an SDNR of 64.3. (b) Matricization applied to a signal-time course from a 1.5T dataset, giving an SDNR of 128.4. (c) Wavelet denoising applied to a signal-time course from a 3T dataset, giving an SDNR of 17.1. (d) Wavelet denoising applied to a signal-time course from a 1.5T dataset, giving an SDNR of 42.4. (e) Tucker decomposition applied to a signal-time course from a 3T dataset, giving an SDNR of 58.8. (f) Tucker decomposition applied to a signal-time course from a 1.5T dataset, giving an SDNR of 189.1.*

Figure 4.12 shows violin plots of the SDNR values from the 3T patient datasets before and after denoising, and Figure 4.13 shows violin plots of the SDNR values from the 1.5T patient datasets before and after denoising, whilst Table 4.4 summarises the mean values ± the standard deviation. All three denoising methods show significantly improved SDNR, with $P < 0.05$, and Tucker decomposition offers significantly better SDNR than matricization and wavelet denoising, with $P < 0.05$, across all the patient datasets.

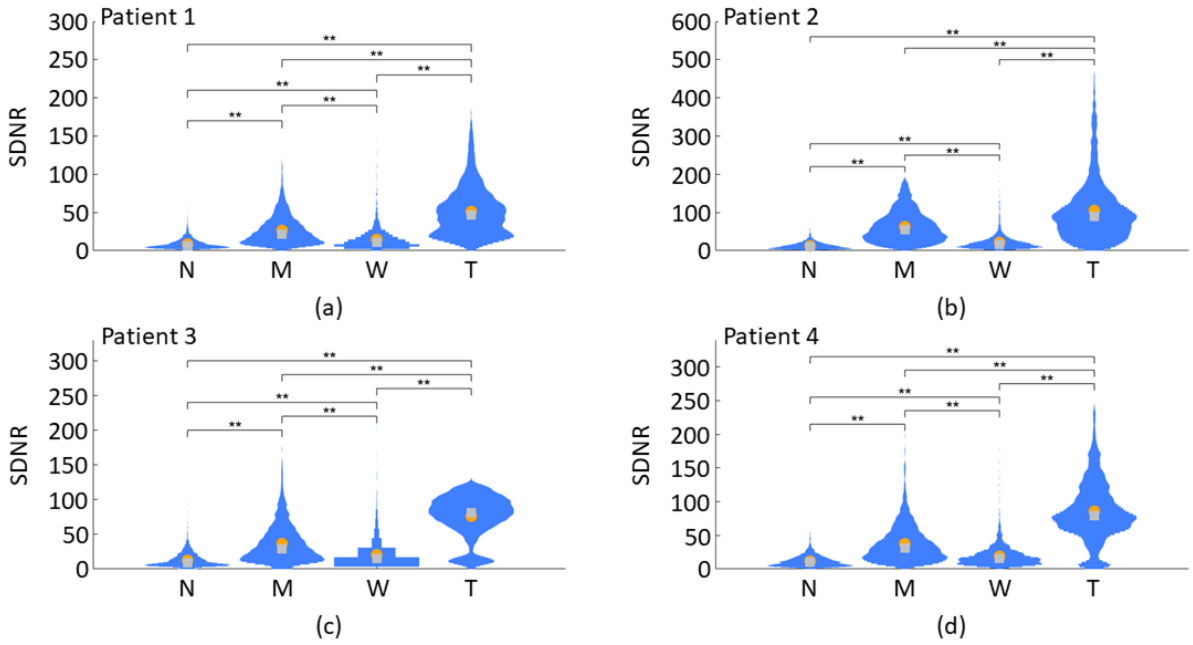



*Figure 4.12: Violin plots of the SDNR values before and after denoising for one slice of each of the 3T patient datasets. Parts a-d correspond to patients 1-4 in Table 4.3. Brackets with * denote P < 0.05, and brackets with ** denote P < 0.001. The orange circles represent the mean value of each distribution, whilst the grey squares represent the median values.*

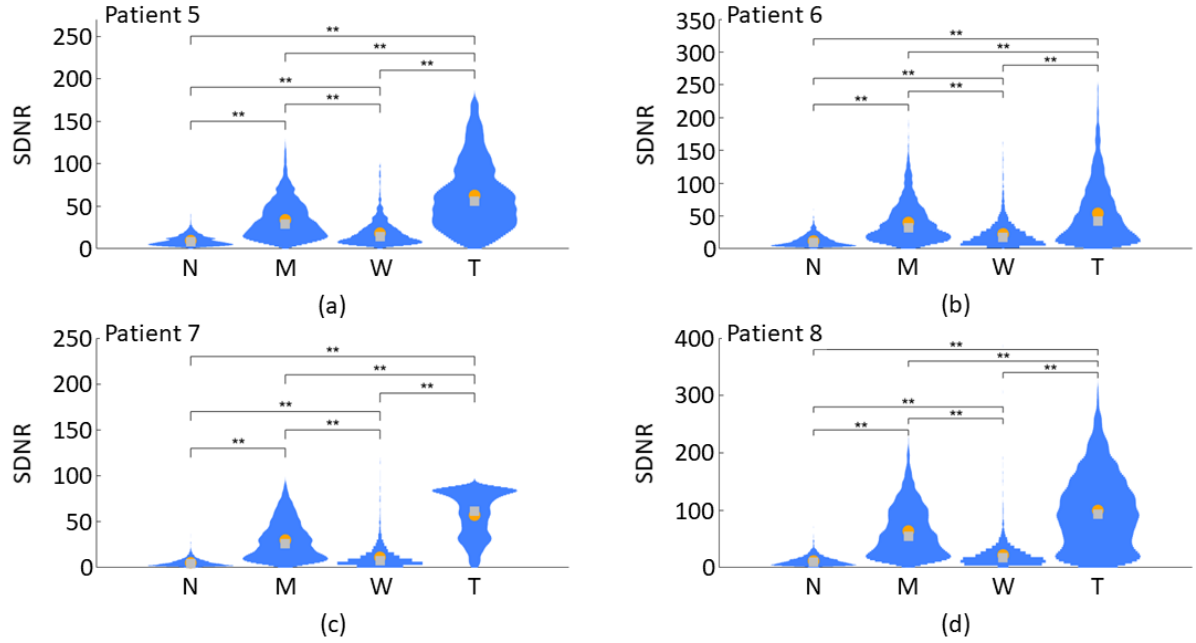



*Figure 4.13: Violin plots of the SDNR values before and after denoising for one slice of each of the 1.5T patient datasets. Parts a-d correspond to patients 5-8 in Table 4.3. Brackets with * denote P < 0.05, and brackets with ** denote P < 0.001. The orange circles represent the mean value of each distribution, whilst the grey squares represent the median values.*

| Patient | Field Strength | Mean SDNR ± Standard Deviation | | | |
|---|---|---|---|---|---|
| | | Pre-Denoising | Matricization | Wavelet Denoising | Tucker Decomposition |
| 1 | 3T | 8.5 ± 6.5 | 26.6 ± 18.9 | 14.7 ± 14.8 | 51.3 ± 32.1 |
| 2 | 3T | 12.4 ± 8.7 | 62.9 ± 39.7 | 21.5 ± 20.0 | 105.3 ± 78.4 |
| 3 | 3T | 12.6 ± 10.0 | 36.8 ± 27.6 | 20.6 ± 32.4 | 76.1 ± 30.0 |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 3T | 11.6 ± 8.1 | 37.5 ± 26.9 | 19.0 ± 15.2 | 85.7 ± 43.2 |
| 5 | 1.5T | 9.5 ± 5.8 | 34.1 ± 21.8 | 18.2 ± 14.2 | 62.7 ± 38.6 |
| 6 | 1.5T | 11.7 ± 8.3 | 40.0 ± 29.8 | 22.6 ± 20.3 | 53.9 ± 44.6 |
| 7 | 1.5T | 5.2 ± 4.0 | 30.0 ± 20.4 | 11.2 ± 13.4 | 57.2 ± 24.8 |
| 8 | 1.5T | 11.5 ± 8.0 | 63.8 ± 42.9 | 22.0 ± 23.0 | 99.5 ± 63.3 |

*Table 4.4: Table showing the mean SDNR ± standard deviation before and after denoising for all patient datasets.*

Figures 4.14 and 4.15 show the rCBV maps produced pre- and post-denoising for the 3T and 1.5T datasets, respectively.

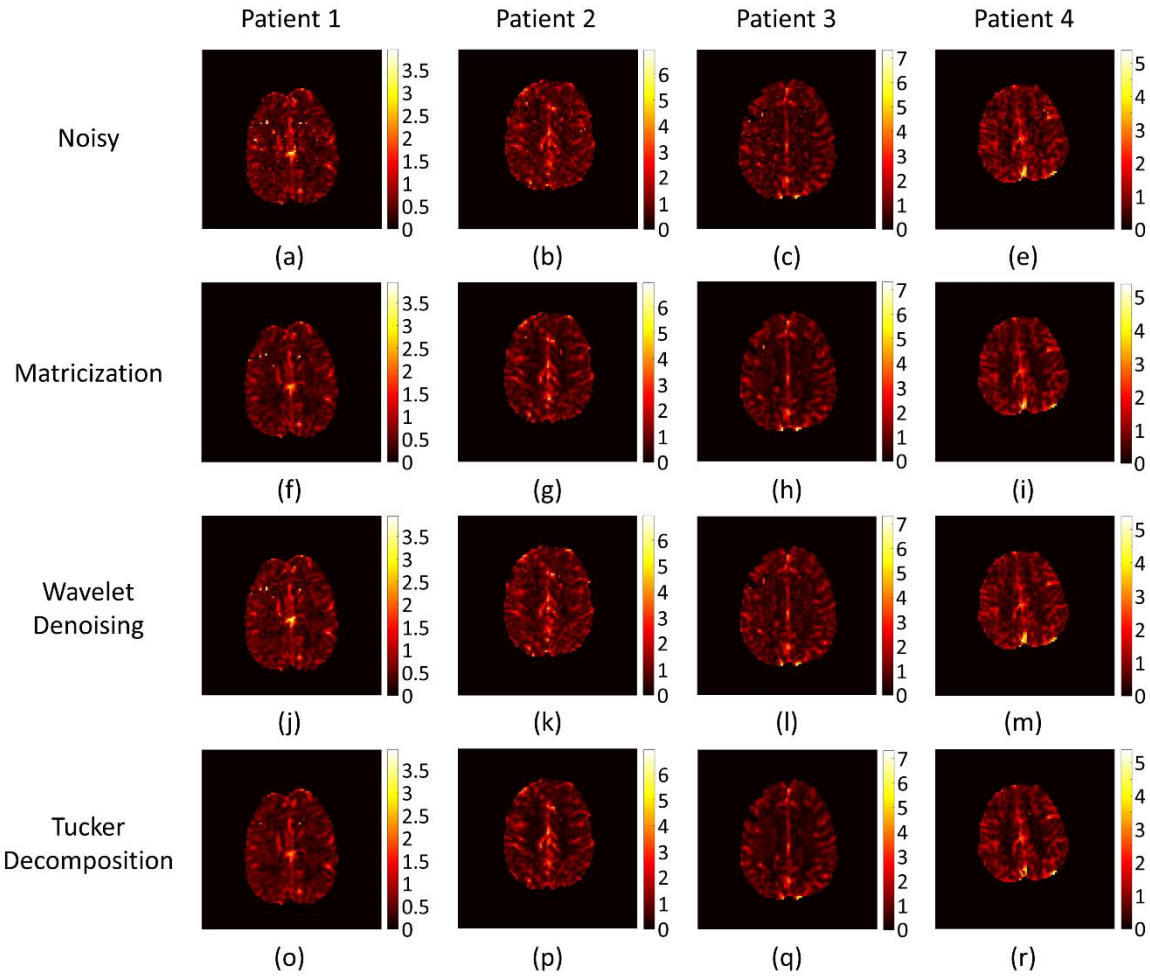



*Figure 4.14: The rCBV maps from the 3T patient datasets pre- and post-denoising. (a-e) shows the rCBV maps pre-denoising for the four 3T datasets; (f-i) shows the rCBV maps post matricization denoising for the four 3T datasets; (j-m) shows the rCBV maps post wavelet denoising for the four 3T datasets; and (o-r) shows the rCBV maps post Tucker decomposition denoising for the four 3T datasets.*

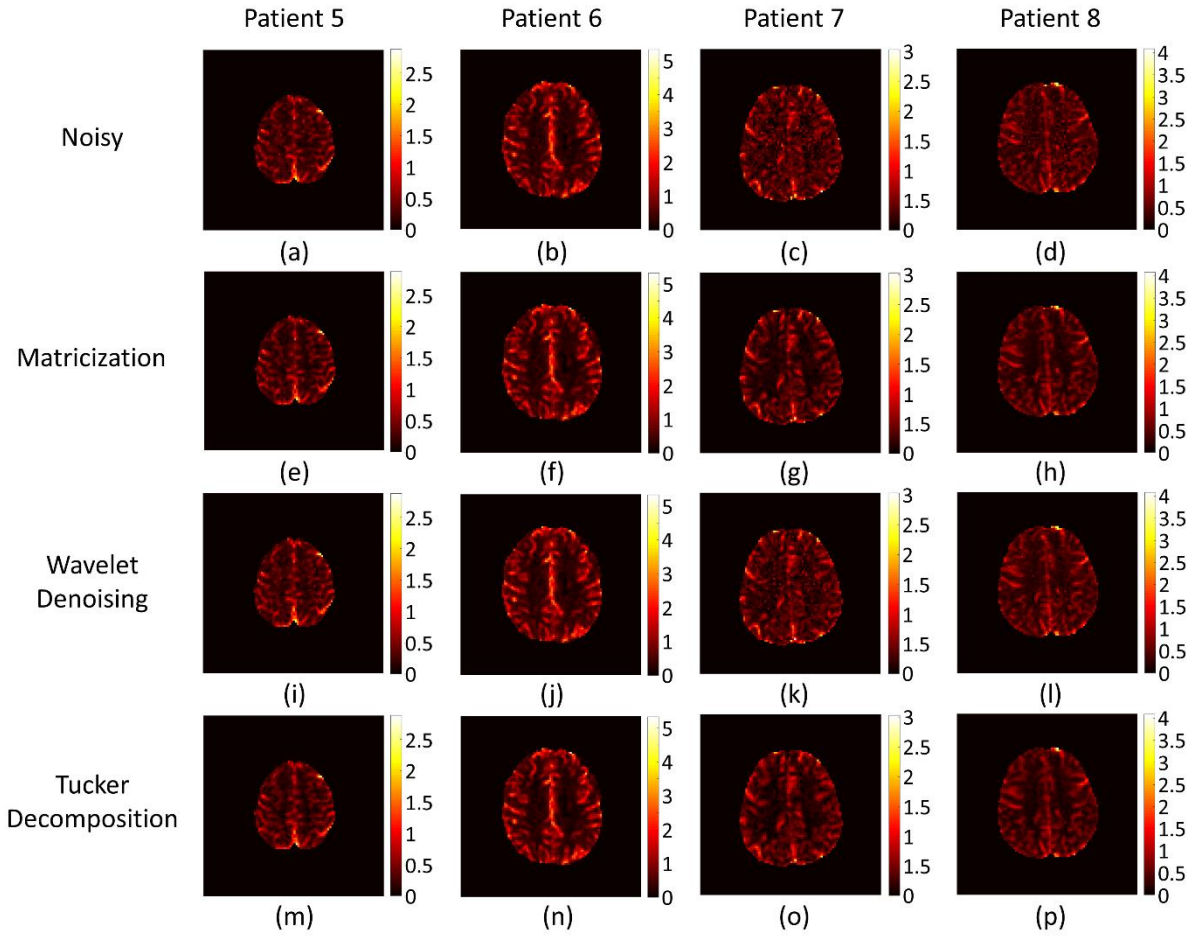



*Figure 4.15: The rCBV maps from the 1.5T patient datasets pre- and post-denoising. (a-e) shows the rCBV maps pre-denoising for the four 1.5T datasets; (f-i) shows the rCBV maps post matricization denoising for the four 1.5T datasets; (j-m) shows the rCBV maps post wavelet denoising for the four 1.5T datasets; and (o-r) shows the rCBV maps post Tucker decomposition denoising for the four 1.5T datasets.*

## 4.4. Discussion

This work shows that denoising can be used to effectively suppress the noise in DSC-MRI signal-time courses in both simulated and patient data. When applied to the simple model, all three methods offer improvements in CBV error across all SDNR values. When applied to the brain models, Tucker decomposition denoising offers improved SDNR compared to matricization and wavelet denoising, but only offers improved CBV error and $\chi^2$ values at low SDNR values. Wavelet denoising offers improved SDNR and $\chi^2$ but no significant improvement in CBV error. In patient data, Tucker decomposition denoising offers improved SDNR compared to matricization and wavelet denoising.

Determining a performance threshold in SDNR is difficult, as figure 4.2 shows that there is no range of SDNR values where there is no improvement in CBV error. Instead, improvement and worsening of CBV error occurs across all SDNR values. All three methods show a majority of signal-time courses have improved CBV error post-denoising and the final column of Table 4.1 shows that averaging across the change in CBV error for every signal-time course shows an overall decrease in CBV error for each denoising method. For wavelet denoising the overall decrease in CBV error is much smaller than the matricization and Tucker decomposition. This is because for wavelet denoising the average increase in CBV error in signal-time courses with a worse CBV error post-denoising is larger than the average decrease in CBV error in signal-time courses with improved CBV error post-denoising.

When applied to the brain models, matricization and Tucker decomposition denoising methods produce good reconstructions of the original signal-time courses, across all three brain models, and offer improved SDNR when compared to noisy data. However, Tucker decomposition reduces noise more effectively at lower SDNR as can be seen from comparing parts (d) and (h) from Figures 4.3 - 4.5. Wavelet denoising removes noise and offers improved SDNR but produces distorted signal-time courses that do not always provide good representations of the original signal. This may be due to fundamental differences in the way each method denoises the data. Wavelet denoising uses the wavelet transform to separate the signal from noise using frequencies. Effectively, the first pass of the signal-time course can be considered to be a very low frequency signal and the noise to be a high frequency signal. As the user has to select the type of wavelet, the number of levels in the decomposition and the thresholding method, the performance of this method is dependent on the assumptions made by the user. In contrast matricization and the Tucker decomposition decompose the signal into components and then separate signal from noise based on variance of each component.

Variance based methods appear to separate signal and noise better than frequency based methods. The variance based method is also data driven and relies less on assumptions by the user. The user still has to select the method to choose the number of components selected by the method. However table 4.2 shows that the number of components for both the matricization and Tucker decomposition methods does not change much with the level of noise in the dataset.

The effect of the denoising methods on the rCBV values in simulated and patient data can be seen in Figures 4.9, and 4.14 – 4.15, respectively. In both the simulated and patient data, the rCBV maps produced by the wavelet denoising show little change from the noisy rCBV maps. When applied to the simulated data, both the matricization and Tucker decomposition methods the rCBV maps look less 'speckled' and there are a lot less failed rCBV fits (represented by black pixels). For some of the patient data sets the matricization and Tucker decomposition produce rCBV maps with less outlier values, and a clearer definition between the GM and WM. For some of the patient data sets the denoising methods do not seem to affect the rCBV maps, but in these cases there is already a clear definition between the GM and WM, so the data is likely to already be good quality prior to denoising.

The violin plots in Figures 4.6 - 4.8 show that both matricization and Tucker decomposition denoising methods give significantly better SDNR, $\chi^2$ and CBV error than noisy data, with $P < 0.05$. Wavelet denoising offers significant improvements in SDNR and $\chi^2$ ($P < 0.05$) but offers no improvement in CBV error ($P = 1$). Tucker decomposition gives significantly better SDNR than matricization across all three brain models with $P < 0.05$. This result agrees with Brender et al., which showed that Tucker decomposition gave larger SDNR than matricization when applied to magnetic resonance spectroscopy (216). Tucker decomposition also changes the shape of the distribution of SDNR values post-denoising, producing a distribution with

two peaks. Matricization and wavelet denoising improve the average SDNR without changing the overall shape of its distribution.

When comparing $\chi^2$ and CBV error values from the violin plots, Tucker decomposition only outperforms matricization for the brain model with an average SDNR value of 4, with $P < 0.05$ for both $\chi^2$ and CBV error. For the brain models with average SDNR values of 8 and 12, matricization gives significantly better $\chi^2$ and CBV error ($P < 0.05$) and for the brain model with an average SDNR of 12, the CBV error for Tucker decomposition is worse than the noisy data ($P < 0.05$). This result shows that although Tucker decomposition performs better at lower SDNR values, matricization is capable of denoising over a larger range of values. This may be due to fundamental differences in the way matricization and Tucker decomposition denoise data. As the Tucker decomposition is applied to a higher dimensional tensor than matricization, it uses a more complex basis to do this. This means that the core tensor can be represented in a more compact form, which results in greater denoising performance (216, 342). However, in the case of the simulated signal-time courses with higher SDNR a small part of the first pass is removed along with the noise, resulting in the first pass being underestimated. Figure 4.10 demonstrates that this is caused by Tucker decomposition underestimating the maximum signal decrease in the first pass as well as removing the noise from higher SDNR signal-time courses, thereby producing a signal-time course with exceptional SDNR but worse CBV error than the noisy signal-time course. This does not appear to occur for all signal-time courses as the average $\chi^2$ and CBV error for Tucker decomposition is significantly better than noisy data ($P < 0.05$).

The standard deviations in the % CBV error values in Table 4.2 are almost as large as the mean values. The brain model contained an equal mix of GM and WM signal-time courses. These signal-time courses have the same amount of noise added to them but have different

sized signal drops and different baseline signal intensities, so are affected by the same amount of noise to a differing extent, leading to a large range in % CBV error.

When applied to the patient datasets, both matricization and Tucker decomposition denoising methods provide good reconstructions of the original signal-time courses for both the 1.5T and 3T datasets, whilst wavelet denoising produces distorted signal-time courses. The violin plots in Figure 4.12 - 4.13 and the mean SDNR values show that Tucker decomposition consistently gives better SDNR than matricization and wavelet denoising, $P < 0.05$. This mirrors the results of the simulations. In the patient datasets used in this chapter the 1.5T data has a larger SDNR value than the 3T data. This may be because the 1.5T patient data was recorded using a single-bolus protocol and the 3T data was recorded using a split-bolus protocol. In paediatrics a split bolus protocol means that half the volume of contrast agent is injected during the DSC acquisition, which leads to a reduction in signal drop, and may lead to reduced SDNR. In the violin plots for the 3T patient datasets the maximum SDNR value on the y axis of the violin plot for patient 2 is double that of the other violin plots. There is no difference in the acquisition protocol in the 3T patient datasets so this may be caused by a few signal-time courses having a larger SDNR due to the Tucker decomposition.

Using simulations has advantages as it allows for post-processing methods to be tested and optimised before they are applied to patient data. For denoising, using simulations allows the improvement in CBV accuracy to be assessed, as the ‘true’ value for CBV is known, which is not possible in patient data. Using a Gaussian distribution of CBV values for simulating signal-time courses adds some variation to the GM and WM signal-time courses, ensuring that they are not all identical and are more representative of a patient dataset. For the simulations in this chapter the SDNR values are much lower than the average SDNR values of the patient datasets. This is because the work shown in Chapter 3 to produce thresholds

from the quantitative measures of quality produced an SDNR threshold of 7.6. The aim of the denoising is to improve the quality of signal-time courses which would normally be classed as poor quality. Therefore, the simulations were given SDNR values similar to the SDNR threshold as these are the signal-time courses that need to be improved by denoising.

Matricization has been applied as a denoising method to multiple types of data previously (211, 214, 216, 326). Therefore, multiple methods for selecting the correct number of components used to reconstruct the data have been developed. Some methods are applicable to multiple types of data. For example, Gurney-Champion et al. have used the components which contained at least 97% of the signal information to denoise diffusion-weighted MRI (201). Ulfarsson et al. present a method for denoising which uses Stein's Unbiased Risk Estimator (SURE) to minimise the mean square error (202). Therefore, selecting the correct method to optimise the number of components used can be challenging.

There are several examples of applying denoising methods to perfusion MRI data, some using the techniques presented in this chapter and some using other techniques. For example Wirestam et al implemented wavelet denoising into the deconvolution of DSC-MRI signal-time courses to improve the quality of the CBF maps produced (324). This differs to the results achieved in this chapter and may be due to the application of a Wiener filter in the wavelet space. An example of another denoising method applied to DSC-MRI data is work by Murase et al., which uses anisotropic diffusion as a denoising method (325). This method was also applied to the deconvolution step and was used to improve the accuracy of CBF values. A third example is the work by Benou et al., which uses deep neural networks to denoise DCE-MRI data, which can be used to improve the accuracy of the parameters obtained (343).

Wavelet denoising has multiple elements that can be adjusted and selected by the user, including the number of levels of denoising applied and the type of wavelet used. This makes selecting the optimal settings challenging. In this chapter the same wavelet and level of denoising has been applied to all signal-time courses. It may be possible to improve the denoising results by selecting individual wavelets and levels of denoising for each signal-time course. However, this would make the denoising process much more manual and would be significantly slower than applying matricization or Tucker decomposition. It may also be possible to improve the method used to optimise the wavelet denoising. However, optimising wavelet denoising is a complex process, whilst matricization and Tucker decomposition offer a simpler, data-driven alternative.

The work presented in this chapter has several limitations. Firstly, the simulations only consider noise in signal-time courses simulated to resemble healthy brain. DSC-MRI is prone to other artefacts such as motion and susceptibility and disease specific artefacts such as contrast agent leakage. This means that real patient data is more complex than the simulation presented here. However, methods to correct other common artefacts found in DSC-MRI have already been implemented. For example, rigid-body registration has been used to correct motion in simulated and patient data, whilst susceptibility artefacts have been corrected for using displacement maps, and leakage correction methods have been developed to account for contrast agent leakage (156, 172, 344). Secondly, there are more complex versions of DSC-MRI models available, such as the model by Mehndiratta et al., which includes a term to account for bolus dispersion (345). The simpler model was chosen as it still provides a good estimate of the signal-time courses, and a more complex model is not necessary for denoising simulations. Thirdly, differing methods were used to select the number of components for all three denoising methods. Although this does not allow for a direct comparison between the

two methods, it allows each method to be optimised independently and better demonstrates their applicability as denoising methods. Finally, the model presented in this work uses Gaussian noise. However, noise in MR images is approximated to be Rician in distribution and Rician distributions are used in simulations to prevent negative noise values and potentially negative intensity values in the simulated image (136). This is only an issue when noise is estimated from the background of an image, which is not the case in DSC-MRI simulations, as the noise is estimated from the baseline signal. Gaussian noise is regularly used in publications on simulated DSC-MRI data, and an example of this is the simulation study published by Knutsson et al. (346).

## 4.5. Conclusion

Denoising of DSC-MRI data is an important step in improving data quality and therefore accuracy of perfusion measures. Matricization and Tucker decomposition both give good reconstructions of the simulated and patient signal-time courses, across the range of SDNR values tested. Both offer an improvement in data quality compared to noisy data. Wavelet denoising produces distorted signal-time courses with no significant improvement in CBV error. Tucker decomposition offers better performance at lower SDNR, but this comes at the cost of reduced performance at higher SDNR values.

# Chapter 5: An Automated Quality Control Pipeline for Assessing and Improving the Quality of Multicentre DSC-MRI Data

# 5. An Automated Quality Control Pipeline for Assessing and Improving the Quality of Multicentre DSC-MRI Data

## 5.1. Introduction

Studies utilising single centre DSC-MRI paediatric data often have small populations. Therefore, there is a need for more multicentre studies. Obtaining robust biomarkers from multicentre paediatric DSC-MRI data is challenging due to the heterogeneous nature of paediatric patient populations, and variability in the acquisition protocols in DSC-MRI data between centres (61).

Currently quality control is carried out by expert qualitative review, as per the ASFNR recommendations (34). This can be subjective and DSC-MRI datasets contain many signal-time courses, which means it is not possible to assess all the data by qualitative review. Therefore, there is a need for a more automated quality control method. Work in Chapter 3 showed that it is possible to automatically assess data quality using a machine learning classifier trained on quality metrics calculated from the signal-time courses. The data quality of DSC-MRI can also be improved by post-processing methods, such as denoising, which could lead to improved biomarker accuracy (325). Work in Chapter 4 showed that it is possible to improve data quality using Tucker decomposition denoising. These two methods could be combined to create a pipeline to automatically assess and improve data quality for DSC-MRI.

Previous work to assess data quality automatically has included a set of statistical thresholds produced by Akella et al. (307), whilst Tucker decomposition has been previously applied as a denoising method to hyperspectral imaging and in MRS, where it showed a significant improvement in SNR and outperformed other denoising methods (216, 227, 228).

The aim of this chapter is to develop a processing pipeline which will automatically assess the quality of DSC-MRI data and improve data quality. In order to achieve this the objectives are: to implement an automated segmentation method to segment the brain from the ventricles; combine the segmentation method with the denoising and automated quality control methods presented in Chapters 3 and 4, to create an automated pipeline; and to apply this pipeline to determine how variations in acquisition and contrast agent injection protocols can affect the quality of the data acquired.

## 5.2. Methods

### 5.2.1. Patient Data

The patient data used in this study was obtained from a multicentre dataset, summarised in Table 5.1. The patient data is from an imaging study entitled "CNS 2004 10 Functional Imaging of Tumours" (NRES REC ref: 04/MRE04/41). Further details on this study can be found in section 3.2.1 of Chapter 3.

| Group | No. of Data sets | Centre | Scanner Type | Protocol | Pre-Bolus? (Y/N) | Field Strength (T) | TR (ms) | TE (ms) | FA ($^o$) | Voxel Vol. ($mm^3$) | Temporal Resolution (s) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 5 | Birmingham | Siemens Avanto | GE-EPI | N | 1.5 | 1490 | 40 | 20 | 28.80 | 1.48 - 1.50 |
| B | 5 | Birmingham | Siemens Avanto | GE-EPI | Y | 1.5 | 1490 - 1643 | 40 | 20 | 28.80 | 1.50 - 1.65 |
| C | 5 | Birmingham | Phillips Achieva | GE-EPI | N | 3 | 1830 - 1835 | 40 | 20 | 21.88 | 1.95 - 1.97 |
| D | 5 | Birmingham | Phillips Achieva | GE-EPI | Y | 3 | 1830 - 1866 | 40 | 20 | 21.88 | 1.95 - 1.98 |
| E | 5 | Birmingham | Phillips Achieva | sPRESTO | N | 3 | 14 - 15 | 21 - 22 | 7 | 41.42 | 1.40 - 2.05 |
| F | 5 | Liverpool | Phillips Achieva | GE-EPI | Y | 3 | 1335 - 2343 | 40 | 75 | 12.25 | 1.47 - 2.58 |
| G | 5 | Newcastle | Siemens Verio | GE-EPI | N | 3 | 1570 | 29 | 45 | 41.42 | 1.57 - 1.58 |
| H | 5 | Nottingham | Phillips Achieva | GE-EPI | Y | 3 | 1860 - 1866 | 40 | 20 | 21.88 | 1.95 - 1.97 |
| I | 5 | Nottingham | Phillips Intera | sPRESTO | N | 1.5 | 23 - 25 | 16 - 17 | 7 | 11.34 - 41.42 | 1.23 - 1.58 |

*Table 5.1: A summary of the protocols used to acquire the multicentre dataset.*

Patients from Birmingham, Liverpool, and Newcastle were all injected with Dotarem contrast agent (Dotarem, Guerbert, France), whilst patients from Nottingham were injected using either Gadovist (Bayer AG, Leverkusen, Germany) or Magnevist (Bayer AG, Leverkusen, Germany). All patients received a total dosage of 0.1 mmol/kg, using a power injector. Patients who did not receive a pre-bolus were injected using a single bolus protocol and received the total dosage in one injection during the DSC acquisition. Patients who received a pre-bolus were injected using a split bolus protocol and received half the total dosage prior to the DSC acquisition, and the other half of the dosage during the DSC acquisition. For all patients, a volume of up to 10 ml of saline was injected after the total dosage of contrast agent was received. All injections were administered at a rate of 3 ml $s^{-1}$.

### 5.2.2. Automated Segmentation of Background and Ventricles

The aim of the segmentation was to produce a mask of all the pixels within the brain, excluding the ventricles, for each DSC image analysed. This was necessary as the ventricles do not contain any blood, and therefore no useful perfusion information. All segmentation was carried out directly on the DSC images, without using conventional $T_1$ or $T_2^*$ weighted images. This was necessary as with multicentre datasets, there may not be access to all of the conventional imaging. Segmenting directly on the DSC images also allows the segmentation to be a fully automated step in the pipeline.

The segmentation was carried out in two stages: the first stage removed the background pixels with the segmentation carried out using the pixel intensities from the first dynamic; whilst the second stage removed the ventricles with the segmentation carried out using features extracted from the signal-time courses.

To remove the background, all of the pixel intensities from the first dynamic of the middle slice of each dataset were separated into two clusters, using k-means clustering, where one

cluster represented the background pixels, and the other cluster represented the brain pixels. The lowest pixel intensity from the brain pixel cluster and the largest pixel intensity from the background pixel cluster were selected. The mean of these two values was calculated and this was applied as a threshold to remove the background pixels from all slices within the dataset.

To remove the ventricles, segmentation was carried out on the signal-time courses, with the segmentation method dependent on the acquisition protocol for the DSC-MRI data. Different methods for segmenting the ventricles were needed as there were differences in the signal-time courses and the contrast of the images produced by the two different protocols.

For GE-EPI data the ventricles were segmented using a combination of PCA and hierarchical clustering applied to the signal-time courses. PCA reduces the dimensionality of a dataset, which makes it possible to identify differences between brain matter and ventricle signal-time courses, whilst ignoring factors such as noise. All signal-time courses selected by the mask during the first stage of segmentation were stacked in a matrix and PCA was applied. Hierarchical clustering was then applied to the scores from the first two principal components and the clustering results were used to separate them into two clusters: one representing the brain matter signal-time courses and one representing the ventricle signal-time courses.

sPRESTO data has less contrast between the brain matter and the ventricles, which meant that the PCA and hierarchical clustering method did not work for this data. Instead the first ratio was used, which is the first dynamic of the signal-time course normalised to the average baseline of the signal-time course (347). For sPRESTO signal-time courses the baseline takes several dynamics to stabilise, which means there is a signal drop between the first dynamic and the stabilised baseline. This signal drop differs between brain matter and ventricle signal-time courses and can be exploited using the first ratio. The first ratio values were calculated

for all signal-time courses selected during the first stage of the segmentation and were then separated into two clusters using k-means clustering. As with the GE-EPI method, one cluster represented the brain matter signal-time courses, and one represented the ventricle signal-time courses.

For two of the Birmingham datasets (one GE-EPI dataset and one sPRESTO dataset) a high resolution $T_2$ weighted image was acquired with the same field of view (FOV) and same number of slices as the DSC dataset. This meant it was possible to segment the brain from the ventricles on the $T_2$ image and register the resulting segmentation to the DSC dataset. The functional magnetic resonance imaging of the brain (FMRIB) group in Oxford have produced a series of tools for processing MRI data, which have been published as the FMRIB software library (FSL) (348). The segmentation of the $T_2$ weighted image was carried out using the FMRIB's automated segmentation tool (FAST) (349). The resulting mask was then imported into Matlab and was manually registered to the DSC-MRI image.

This mask was used as a 'gold standard' comparison to the automated segmentation method. Birmingham datasets were used because $T_2$ weighted images with the same FOV and same number of slices as the DSC were not always available from other centres. The two segmentation methods were compared using the modified Hausdorff distance (350) and the dice similarity coefficient (351).

### 5.2.3. Denoising

The aim of the denoising step was to improve the quality of the data. This was carried out using the Tucker decomposition method presented in Chapter 4. This method was chosen as it produced the best results at low SDNR and these are the signal-time courses which are most likely to fail quality control. The Tucker decomposition method is also well suited to denoising DSC-MRI datasets as it is specifically designed to be applied to higher dimensional

data (216). It was concluded that the improved performance at lower SDNR values outweighed the possible risk of a slight reduction in data quality at higher SNR as the actual increase in CBV error in the simulated data presented in Chapter 4 was small.

As in Chapter 4, the Tucker decomposition was applied using the N-way (221) and tensorlab (341) toolboxes in Matlab. In this chapter entire patient datasets are being denoised instead of individual slices. Therefore, each patient dataset was treated as a four way tensor, with the first two dimensions being the rows and columns of the image, the third dimension being the slices and the fourth dimension being the signal-time courses. The mlrankest function from the tensorlab toolbox was used to estimate the rank of the core tensor and the result from this was used as the number of components used in the Tucker decomposition.

### 5.2.4. Calculation of Metrics

For each signal-time course, SDNR, RMSE, FWHM and PSR were calculated as before. Pixel-by-pixel rCBV was also calculated by computing the area under a gamma variate fit to the first pass of the CTC. All rCBV values were normalised to average rCBV from a WM mask. The method used to produce the WM mask is described in the next section. Further details on how each of the metrics were calculated can be found in Chapter 2, and details on how the baseline and post-bolus signal was identified can be found in section 3.2.3. of Chapter 3.

#### 5.2.4.1. Normalisation of rCBV

In order to normalise the rCBV maps produced by the pipeline, it was necessary to segment WM from the DSC-MRI images, to produce an average rCBV for normally appearing WM (NAWM). WM segmentation was achieved using a network created in MeVisLab (MeVis Medical Solutions AG, Bremen, Germany, version 2.8.2). This network used T2 weighted images to segment the WM and then register the resulting segmentation to the DSC image.

The first step was to upsample the DSC image, which was achieved using nearest neighbour resampling (352). The T2 image was reformatted and registered to the upsampled DSC image to ensure that it had the same FOV as the DSC image. Following this the reformatted T2 image was segmented using region growing methods. In this method one or more seed points and a threshold value are selected, and any pixels that fall within the threshold of the average of the seed points are segmented. The resulting mask is then downsampled to the dimensions of the original DSC image.

#### 5.2.5. Assessing Data Quality

Quality control was carried out using a random forest machine learning classifier from the machine learning toolbox in Matlab. This machine learning classifier was selected as it produced the lowest classification error when compared to other machine learning classifiers in Chapter 3. The classifier was trained using 1,027 signal-time courses from a multicentre dataset. SDNR, RMSE, FWHM and PSR were used as predictors of quality. More details on the machine learning can be found in Chapter 4.

#### 5.2.6. Statistical Analysis

Statistical testing was used to assess differences in the quality metrics and % passing quality control before and after denoising and between centres. Kruskal-Wallis testing was used to investigate differences in quality metrics and $\chi^2$ testing was used to investigate differences in the % of signal-time courses passing quality control. All statistical testing was carried out using R (R Foundation for Statistical Computing, Vienna, Version 3.5.0).

### 5.3. Results

#### 5.3.1. Constructing the Pipeline

Figure 5.1 summarises the processes involved in each stage of the pipeline. Figure 5.2 shows an example of the differences between a GE-EPI and sPRESTO signal-time course, which were both acquired with a single-bolus protocol. Figure 5.3 shows some examples of the

automated segmentation method used in the pipeline compared to the standard method of segmentation. Table 5.2 summarises the results from the performance metrics used to compare the two segmentation methods. Figure 5.4 shows an example of the WM masks produced by the network in Mevislab.

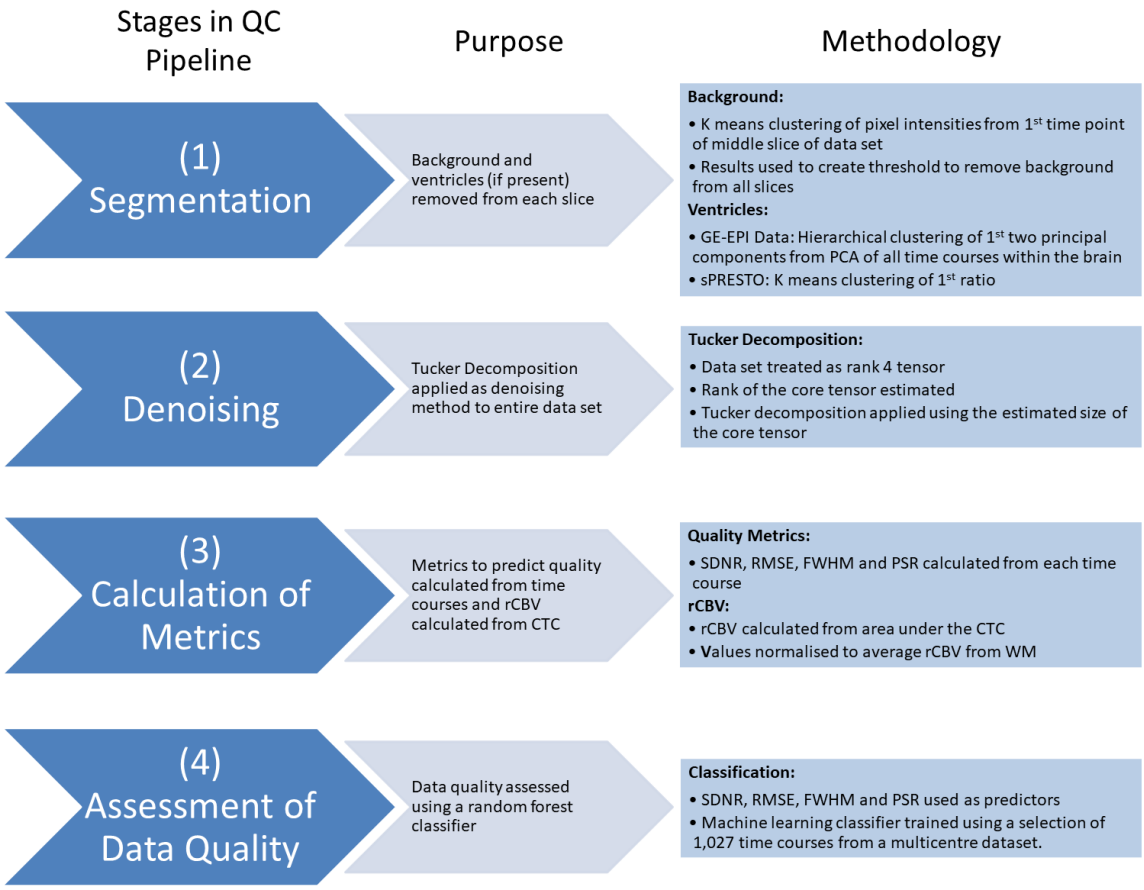



*Figure 5.1: Flowchart summarising the various processes within the quality control pipeline.*

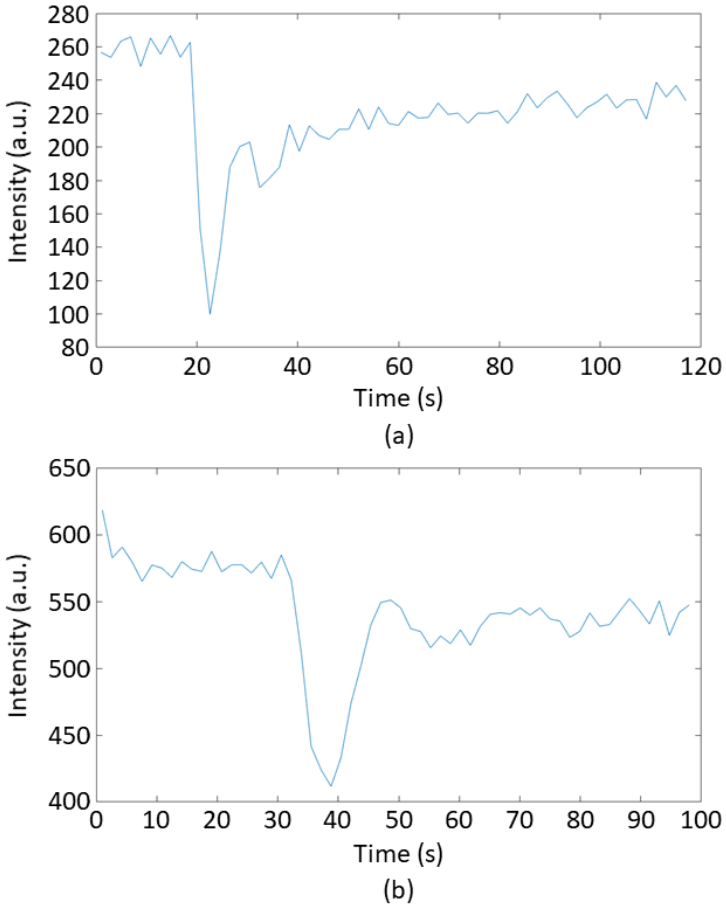



*Figure 5.2: An example of the signal-time courses from (a) a GE-EPI dataset and (b) an sPRESTO dataset.*

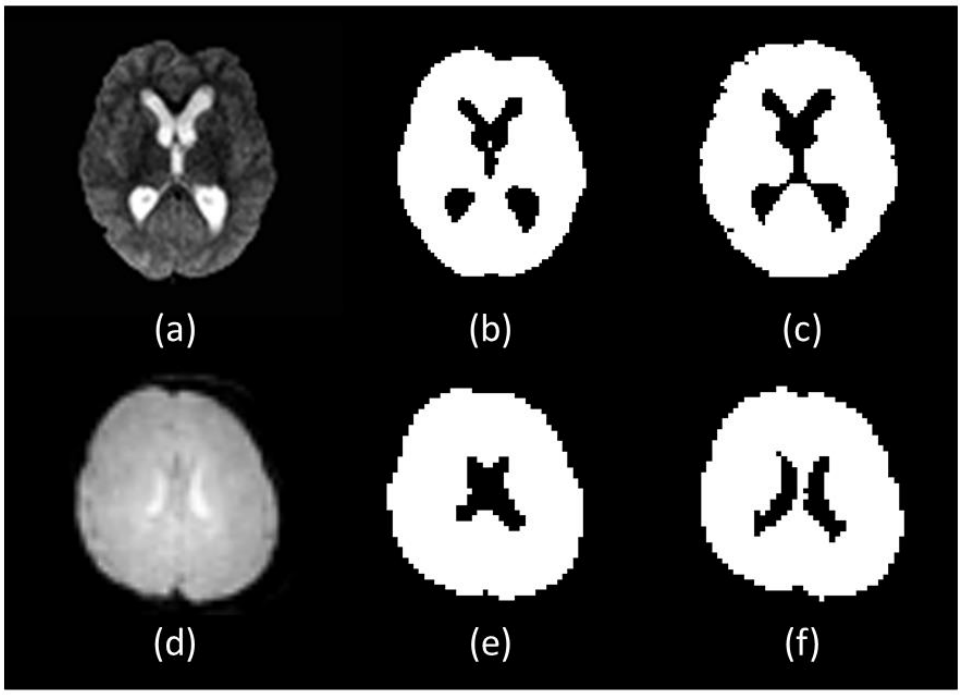



*Figure 5.3: Examples of the ventricle segmentation using the automated method developed (middle column) compared to applying FSL to a $T_2$ weighted image registered to the DSC-MRI scan (right column), for an GE-EPI (top row) and sPRESTO acquisition (bottom row). (a) is a slice from a GE-EPI DSC-MRI dataset, (b) is the automated segmentation applied to this*

*image, and (c) is FSL applied to the $T_2$ weighted image and registered to the DSC image. (d) is a slice from a sPRESTO DSC-MRI dataset, (e) is the automated method applied to this image, and (f) is FSL applied to a $T_2$ weighted image and registered to the DSC image.*

| Protocol | Modified Hausdorff Distance (mm) | Dice Similarity Coefficient | Pixel Spacing (mm) | Slice Thickness (mm) |
|---|---|---|---|---|
| GE-EPI | 1.71 | 0.95 | 2.39 | 3.5 |
| sPRESTO | 3.44 | 0.92 | 3.43 | 3.5 |

*Table 5.2: Performance measures comparing the automated segmentation method to FSL of a $T_2$ weighted image registered to the DSC image.*

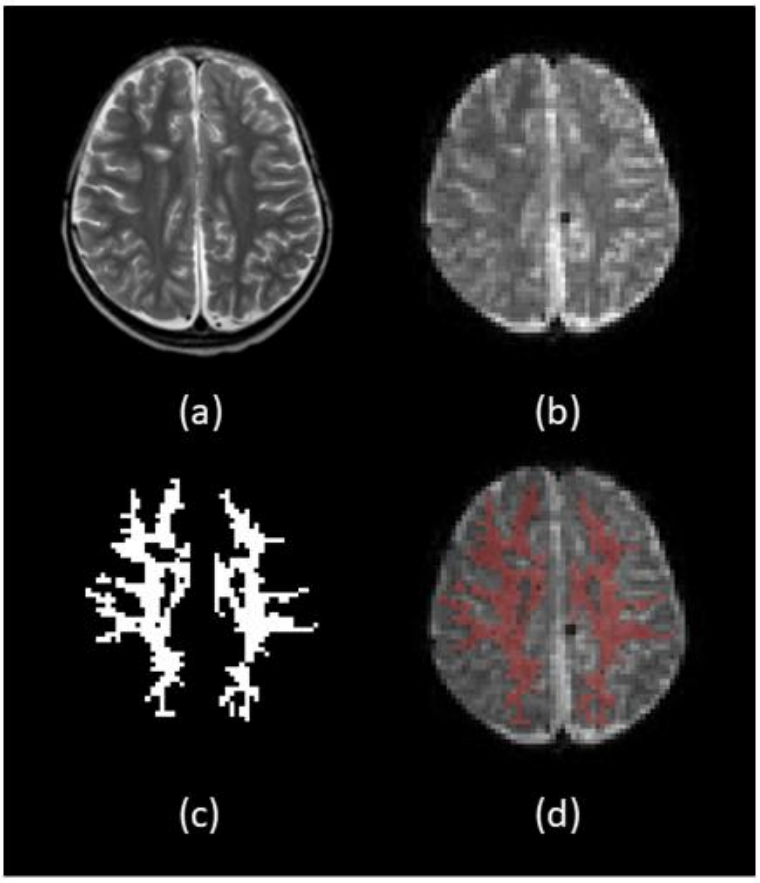



*Figure 5.4: An example of the WM segmentation produced by the Mevislab network. (a) is a $T_2$ weighted image, (b) is the original DSC image, (c) is the WM mask produced and (d) is the WM mask overlaid onto the DSC image.*

### 5.3.2. Outputs from the Pipeline

Figure 5.5 shows an example of the maps produced by the pipeline for one slice of a 3T GE-EPI dataset with a pre-bolus from Birmingham. Table 5.3 summarises the average metrics on a centre-by-centre basis pre- and post-denoising. Figure 5.6 shows bar charts showing the changes in average SDNR and % passing for each centre pre- and post-denoising.

Out of all the datasets included in the study, Kruskal-Wallis testing showed that there was a significant improvement ($P < 0.05$) in SDNR and RMSE post-denoising for 84.4% and 73.3% of datasets, respectively. Chi square testing showed there was a significant improvement ($P < 0.05$) in the % of signal-time courses passing quality control post-denoising for 73.3% of datasets. Kruskal-Wallis testing gave $P < 0.01$ when comparing all metrics between centres pre- and post-denoising. Chi square testing gave $P < 0.01$ when comparing the % of signal-time courses passing quality control between centres.

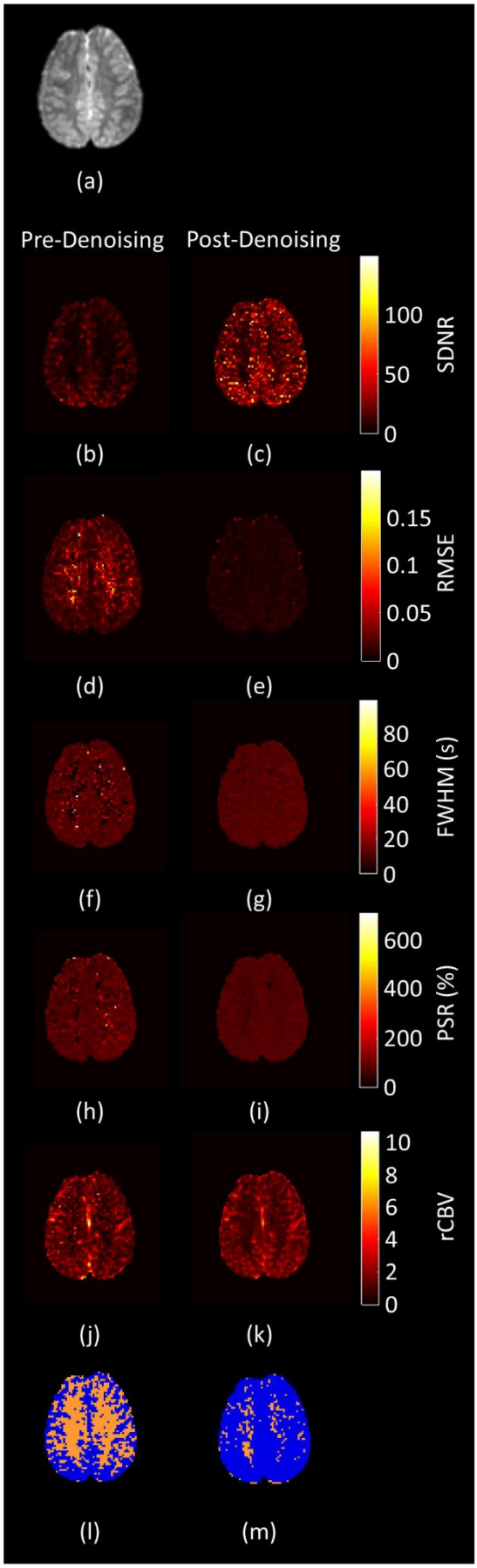

(a)
Pre-Denoising
Post-Denoising
100
50
0
SDNR
(b)
(c)
0.15
0.1
0.05
0
RMSE
(d)
(e)
80
60
40
20
0
FWHM (s)
(f)
(g)
600
400
200
0
PSR (%)
(h)
(i)
10
8
6
4
2
0
rCBV
(j)
(k)
(l)
(m)

*Figure 5.5: An example of the maps produced by the quality control pipeline, for one slice of a GE-EPI DSC-MRI dataset. (a) shows the original DSC image, (b) and (c) show the SDNR maps pre- and post-denoising, (d) and (e) show the RMSE maps pre- and post-denoising, (f) and (g) show the FWHM maps pre- and post-denoising, (h) and (i) show the PSR maps pre- and post-denoising, (j) and (k) show the rCBV maps pre- and post-denoising, and (l) and (m) show the quality control maps pre- and post- denoising. For the quality control maps, blue pixels represent signal-time courses that passed quality control, whilst orange pixels represent signal-time courses which failed quality control.*

| Group | Average SDNR ± Std. Dev. | | Average RMSE ± Std. Dev. | | Average FWHM ± Std. Dev. | | Average PSR ± Std. Dev. (%) | | Average % Passed ± Std. Dev. (%) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post | Pre | Post | Pre | Post |
| A | 8.7 ± 6.9 | 144.4 ± 133.5 | 0.035 ± 0.560 | 0.008 ± 0.062 | 7.8 ± 5.1 | 8.3 ± 2.3 | 80.8 ± 13.9 | 81.1 ± 10.6 | 51.3 ± 21.4 | 95.9 ± 5.9 |
| B | 3.9 ± 4.6 | 57.6 ± 99.7 | 0.111 ± 3.172 | 0.022 ± 0.527 | 7.8 ± 8.8 | 9.0 ± 5.0 | 78.5 ± 22.7 | 78.4 ± 18.0 | 18.2 ± 23.8 | 76.7 ± 28.0 |
| C | 15.6 ± 10.9 | 37.5 ± 43.1 | 0.025 ± 0.091 | 0.022 ± 0.045 | 9.7 ± 6.0 | 9.6 ± 5.4 | 73.9 ± 12.9 | 73.8 ± 12.6 | 77.0 ± 10.2 | 84.8 ± 15.3 |
| D | 8.4 ± 8.5 | 52.9 ± 68.1 | 0.043 ± 0.601 | 0.020 ± 0.104 | 10.4 ± 8.3 | 10.9 ± 5.6 | 71.0 ± 22.6 | 68.9 ± 19.9 | 41.6 ± 21.9 | 72.4 ± 37.0 |
| E | 23.1 ± 16.3 | 66.4 ± 66.1 | 0.016 ± 0.023 | 0.016 ± 0.097 | 7.0 ± 3.6 | 6.7 ± 2.4 | 81.3 ± 10.6 | 82.1 ± 9.9 | 92.3 ± 4.1 | 94.4 ± 5.2 |
| F | 26.1 ± 16.7 | 139.8 ± 110.9 | 0.015 ± 0.339 | 0.008 ± 0.148 | 10.0 ± 3.7 | 10.3 ± 3.3 | 74.9 ± 14.9 | 74.6 ± 15.3 | 91.6 ± 10.7 | 98.4 ± 1.7 |
| G | 47.8 ± 30.1 | 51.5 ± 45.4 | 0.009 ± 0.011 | 0.006 ± 0.006 | 9.8 ± 2.8 | 10.7 ± 2.4 | 77.1 ± 10.9 | 76.2 ± 9.0 | 96.4 ± 1.6 | 96.8 ± 1.5 |
| H | 14.3 ± 11.0 | 19.9 ± 17.9 | 0.023 ± 0.223 | 0.020 ± 0.147 | 8.3 ± 4.1 | 8.4 ± 3.5 | 75.1 ± 16.2 | 74.9 ± 15.6 | 73.6 ± 16.5 | 79.6 ± 18.6 |
| I | 12.5 ± 9.8 | 19.2 ± 12.3 | 0.018 ± 0.128 | 0.018 ± 0.901 | 8.1 ± 4.1 | 8.3 ± 3.5 | 79.3 ± 16.1 | 81.8 ± 15.0 | 75.5 ± 13.7 | 88.6 ± 4.3 |

*Table 5.3: A summary of the centre averages for each of the metrics pre- and post-denoising.*

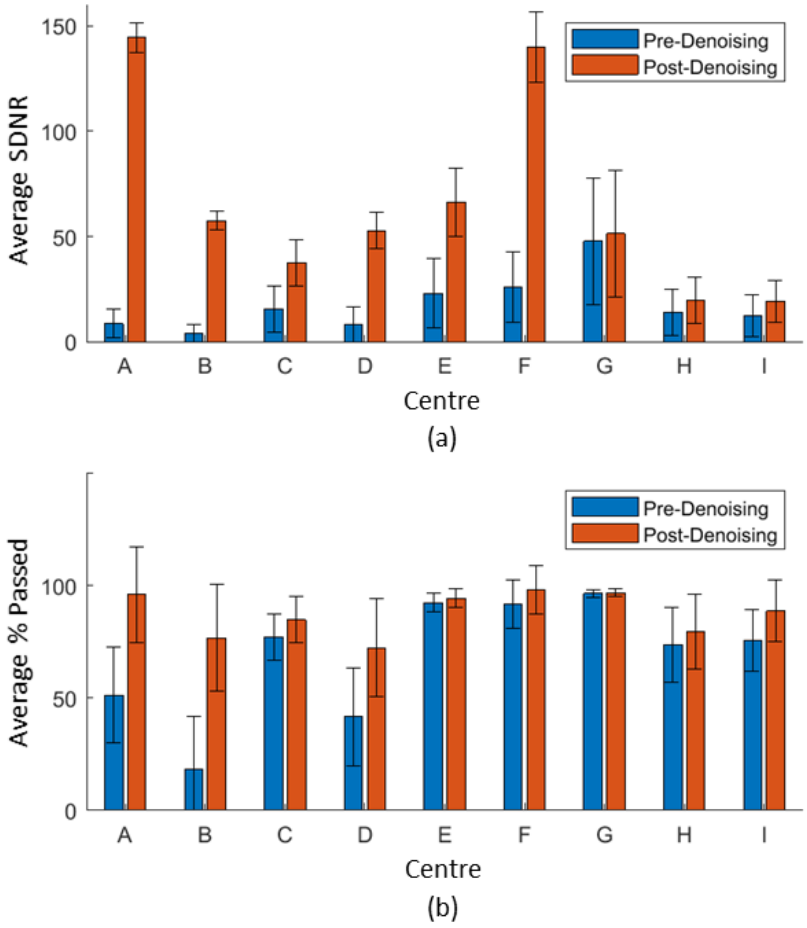



*Figure 5.6: Bar charts showing (a) the change in average SDNR and (b) the change in % passing ± the standard deviation for each centre pre- and post-denoising.*

## 5.4. Discussion

The work in this chapter shows that it is possible to create an automated quality control pipeline, which is capable of segmenting DSC-MRI images, denoising the signal-time courses and assessing data quality using machine learning. Denoising provides a significant improvement in data quality for the majority of datasets. Statistical testing also shows significant differences in data quality between centres, showing the impact of acquisition protocol on data quality.

The use of the automated segmentation method applied directly to the DSC images is advantageous as it does not require any conventional imaging, which may not always be available in a multicentre dataset. It also allows the segmentation to be a fully automated step in the processing pipeline. Comparing the automated segmentation methods for the GE-EPI

and sPRESTO images to the FSL gold standard segmentation shows similar performance, for the two example datasets which were tested. The modified Hausdorff distances for both datasets are less than the pixel spacing and the slice thickness, which shows that for the two datasets the borders were less than a pixel different when compared to the FSL gold standard segmentation. The dice similarity coefficient for both segmentation methods produces values greater than 0.9, showing good agreement with the FSL gold standard segmentation. As this method has only been tested on one example of GE-EPI and one example of sPRESTO data, more datasets would need to be tested before it can be concluded if the new segmentation method is comparable to the gold standard.

Differing segmentation methods were required due to the differences in the image contrast produced by each acquisition protocol. GE-EPI has good contrast between the ventricles and brain matter and PCA of the signal-time courses combined with hierarchical clustering was able to separate brain matter from ventricles. However, this did not work for sPRESTO due to lack of contrast between these regions. Therefore, the first ratio which exploits differences in the signal drop between the first time point and the stabilised baseline between brain matter and ventricles, was applied instead. Comparing the signal-time courses in Figure 5.2, the differences in the baseline signal of GE-EPI and sPRESTO signal-time courses can be seen.

Comparing the example maps shown in Figure 5.5 pre- and post-denoising, an improvement in SDNR, RMSE and % passing quality control can be clearly seen. The change in FWHM, PSR and rCBV are less pronounced but that is because there is a linear relationship between an increased SDNR, decreased RMSE, and increased % of signal-time courses passing quality control. For PSR and FWHM quality metrics there is no linear relationship between data quality and change in PSR or FWHM, as a very low or very high value of PSR or FWHM indicates poor quality data. Therefore, denoising corrects outlier values of PSR and

FWHM. In the case of rCBV, poor quality data makes it difficult to fit a curve to the first pass and leads to underestimated or overestimated rCBV values. Consequently, denoising corrects outlier values of rCBV and there is a non-linear relationship between rCBV and data quality.

This trend is repeated in the centre averages in Table 5.3, with SDNR, RMSE and % of signal-time courses passing quality control showing improvement post-denoising. For SDNR, the standard deviation increases post-denoising, suggesting that the increase in SDNR is not uniform across all datasets. For the rest of the metrics there is a general decrease in the standard deviation post-denoising, suggesting that for metrics such as FWHM, denoising corrects outlier values to a greater extent than those closer to the mean.

The P values produced by comparing the results on a centre-by-centre basis show that the acquisition protocol used to acquire the data has a significant effect on data quality. Several conclusions can be drawn by comparing the results from applying several different acquisition protocols at one centre (Birmingham). Firstly, 3T data is of better quality than 1.5T, which is expected as a larger field strength gives greater SDNR. Secondly, sPRESTO generally gives a higher SDNR and % of signal-time courses passing quality control than EPI. However, sPRESTO has poorer temporal and spatial resolution and poorer contrast than GE-EPI (353). The reduced spatial resolution also means that the sPRESTO sequence has a larger voxel size than GE-EPI, which may be the cause of the increased SDNR compared to GE-EPI. Thirdly, using a single bolus protocol gives a higher SDNR and % of signal-time courses passing quality control than giving a pre-bolus. Comparing the centres that used GE-EPI it can be seen that both the Liverpool and Newcastle datasets gave the largest SDNR. Newcastle data was consistently acquired using a single-bolus protocol and both centres used larger flip angles in their GE-EPI protocols than other centres. Although a larger flip angle leads to greater SDNR, it also leads to increases $T_1$ effects in the acquired image (132).

It can be seen from the bar charts in Figure 5.6 that there is a link between improved SDNR and improved % of signal-time courses passing denoising. For every centre, an improvement in average SDNR leads to an improvement in average % of signal-time courses passing quality control. For centre F, a large improvement in SDNR only leads to a slight improvement in % passing quality control. This is because the % passing quality control was already close to its maximum value pre-denoising.

Inspecting the results on a dataset by dataset basis, it can be seen that significant improvement in data quality is not universal across all datasets, with 73.3% of all the datasets showing a significant improvement in the % of signal-time courses passing quality control post-denoising. However, a larger percentage of datasets had a significant improvement in SDNR post-denoising, showing that a significant improvement in SDNR does not necessarily lead to a significant improvement in data quality. This is because the % of signal-time courses passing quality control has an upper limit, whereas the SDNR value does not. Therefore, for a dataset that is already high quality prior to denoising, a change in SDNR will have a negligible impact in the % of signal-time courses passing quality control.

In terms of other work there have been automated methods to assess data quality, such as the work by Akella et al. (307), which attempted to determine quality thresholds for DSC-MRI. In this work the signal drop of the first pass, PSR, FWHM, and % failure rate of fitting a gamma variate to the first pass were used as quality metrics. Average values were calculated for each patient and a one-sided 99% confidence interval was applied and used to determine a threshold for each of the metrics (307). This differs from the work presented here as it uses different quality metrics, and uses statistical thresholds, instead of thresholds based on qualitative review. However, most other studies seem to rely on the ASFNR recommendations to assess quality by eye prior to processing (34). Denoising has previously been applied to DSC-MRI and other MR modalities such as MRS (216). However, there has

not been any previous work which combines denoising with automated quality control for DSC-MRI.

There are currently several limitations to this work. Firstly, the ventricle segmentation has only been compared to the gold standard segmentation method for a limited number of datasets and may not work well with DSC data acquired with other acquisition protocols. This could be improved by testing comparing the performance of the automated and gold standard methods for a much larger range of datasets. Secondly, the WM segmentation method is currently manual, which prevents the pipeline from being fully automated. This could be improved by implementing automated WM segmentation. Finally, the pipeline and its components have not yet been tested on imaging slices which contain brain pathology, such as tumour. The machine learning classifier has not been trained on tumour data, so may not be able to classify its quality accurately. Tumour signal-time courses may be affected by contrast agent leakage, which may lead to them being misclassified due to abnormal PSR values. However, this could be mitigated for by applying leakage correction and including leakage corrected signal-time courses in the training data. The segmentation method has also not been tested on slices containing brain tumour. Whether it is able to distinguish tumour from ventricle, will likely depend on the type of tumour and the perfusion levels. Contrast agent leakage correction was not applied as the brain slices analysed did not contain brain tumour. Work presented in Appendix 1 showed that leakage had an insignificant effect on the rCBV values obtained.

## 5.5. Conclusion

This work demonstrates that it is possible to create an automated quality control pipeline for DSC-MRI. There are significant differences in data quality caused by variations in acquisition protocol. Tucker decomposition denoising can be used to significantly improve the quality of the majority of datasets, in particular those with low SDNR.

# Chapter 6: Application of the Automated Quality Control Pipeline to Data from Paediatric Brain Tumour Survivors

# 6. Application of the Automated Quality Control Pipeline to Data from Paediatric Brain Tumour Survivors

## 6.1. Introduction

Brain tumours are the main cause of death due to cancer in children (354). They can be challenging to treat, as their location can make them difficult to remove with surgery (355). The three treatments usually applied to brain tumours are surgery, radiotherapy, and chemotherapy. Surgery to resect the tumour, if possible, is usually the first stage of treatment, with a sample of the resected tumour used for histologic diagnosis (354). Radiotherapy may be either craniospinal (CS) or focal. In CS radiotherapy, the entire brain and spinal cord is irradiated, and this is used for malignant tumours, such as medulloblastomas, which have a significant risk of metastasis in the brain and spine (356, 357). In these patients, irradiating the entire brain has the advantage of a reduced chance of metastatic recurrence (358). However, there are concerns over the use of CS radiotherapy in paediatric patients, as this may cause damage to normal brain particularly in the very young, which could lead to problems such as cognitive deficiencies, growth retardation, hormonal deficiency and the potential to cause severe physical disability in the patient (356-358). Focal radiotherapy refers to a group of radiotherapy techniques where the dose of radiation is highly targeted on the brain tumour or region of the brain it arose from, which reduces the dose of radiation to normal brain (358). Examples of focal radiotherapy techniques include focal fractionated radiotherapy (FFRT), stereotactic radiosurgery (SRS) and proton beam therapy (358). Chemotherapy is also used in some tumour types with an intent to increase survival, and is preferred to radiotherapy in young children as it reduces their exposure to ionising radiation (354).

Radiotherapy treatments reduce the CBV within the tumour but can also lead to reduced CBV in normal brain tissues (359-361). Chemotherapy is also used to reduce tumour perfusion and tumours with reduced perfusion are classed as responsive to treatment (362). However, in

some cases treatments may be used to increase blood flow to tumour to aid in the delivery of chemotherapy drugs (363).

DSC-MRI can be used following radiotherapy to assess changes in perfusion in the tumour and in normal brain. This is routinely carried out in adults but is not common in paediatrics. In adults, DSC-MRI has previously been used to distinguish between recurrence and radiation necrosis (364), and to predict survival in high grade glioma before and during radiotherapy (365).

The aim of this chapter was to apply the quality control pipeline developed in the last chapter to DSC-MRI scans acquired after completion of treatment from a cohort of patients with posterior fossa tumours which have undergone surgical resection but differing adjuvant treatment. The effect of different treatments on perfusion in normal appearing supratentorial brain tissue was assessed by comparing the rCBV values between treatment groups.

## 6.2. Methods

### 6.2.1. Patient Data

All patient datasets came from an imaging study entitled "CNS 2004 10 Functional Imaging of Tumours". Further details on this study can be found in section 3.2.1 of Chapter 3. All DSC scans were acquired on a Phillips Achieva 3T scanner at Birmingham Children's Hospital, with TE = 40 ms, TR = 1830-2225 ms, flip angle = 20°, voxel dimensions = 2.5 mm x 2.5 mm x 3.5 mm, scan type = GE-EPI, field of view = 230 mm x 230 mm, number of slices = 30, parallel imaging = SENSE with factor 2, temporal resolution = 1.83 - 2.23s, no. of dynamics = 60, total acquisition time = 117s. The range of TR values is large as one patient scan was recorded with a different head coil to the rest (the SENSE-NV-16 coil), which produced a TR of 2225ms. For the rest of the patients the TR values varied from 1830ms to 1865ms.

All patients were injected with Dotarem contrast agent (Dotarem, Guerbert, France), with a total dosage of 0.1 mmol/kg, using a power injector. Some patients were injected using a single bolus protocol and received the total dosage in one injection at the start of time point 6 in the DSC data acquisition. Some patients were injected using a split bolus protocol and received half the total dosage prior to the DSC acquisition, and the other half of the dosage at time point 6 in the DSC acquisition. This was to minimise $T_1$ weighted effects from contrast agent leakage (as discussed in section 2.2.2.2.). For all patients, a volume of up to 10 ml of saline was injected after the total dosage of contrast agent was received. All injections were administered at a rate of 3 ml $s^{-1}$.

The patients in this study were all diagnosed with cerebellar brain tumours and the tumour types included: 13 Medulloblastomas, 6 Ependymomas, 5 Pilocytic Astrocytomas, and 1 Atypical Choroid Plexus Papilloma. Only patients with cerebellar tumours were included so that it would be possible to compare the effects of treatment on supratentorial slices, which did not include any tumour. All patients initially underwent surgery to remove the tumour, and all scans were acquired more than one year after the completion of all treatment. The patients were split into three groups: those that received CS radiotherapy, those that received focal radiotherapy to the posterior fossa, and those that did not receive radiotherapy. Some of the patients in this study also received chemotherapy. Table 6.1 summarises the number and age range of patients in each treatment group, whilst Table 6.2 summarises the treatments received by each patient.

| Treatment | No. of Patients | Age at Diagnosis (yrs.) | |
|---|---|---|---|
| | | Average | Range |
| Focal Radiotherapy | 4 | 3.3 | 1.4 - 4.1 |
| CS Radiotherapy | 9 | 5.6 | 2.5 - 10.1 |
| No Radiotherapy | 12 | 3.2 | 0.3 - 8.8 |

*Table 6.1: A summary of the number of patients and age ranges included in each treatment group.*

| CCLG ID | Treatment Group | Tumour Type | Age at diagnosis | Chemotherapy | | Radiotherapy | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Date | Details | Date | Details | Dose |
| 1189 | Craniospinal Radiotherapy | Medulloblastoma M2 | 2yrs 6m | T1: Oct 2013 - Mar 2014<br>T2: Apr 2014 | T1: HEADSTART II protocol<br>T2: High Dose chemotherapy and blood stem cell rescue | Sep 2014 - Oct 2014 | Radiotherapy after relapse in Aug 2014. Craniospinal radiotherapy to whole CNS, with boost to posterior fossa. | 35Gy, with boost of 20Gy |
| 1557 | Craniospinal Radiotherapy | Medulloblastoma M3 | 10yrs 1m | Completed May 2016 | POG 9031 chemotherapy | Completed Apr 2016 | Craniospinal tomotherapy with posterior fossa boost | 39.6Gy with 16.2Gy boost to tumour bed |
| 1603 | Craniospinal Radiotherapy | Medulloblastoma M0 | 4yrs 3m | Mar 2016 - Mar 2017 | Standard risk CCLG guidelines - 8 cycles of "Packer" chemotherapy - 4 of (Cisplatin, CCNU, Vincristine), 2 of (carboplatin, CCNU, vincristine), 2 of (CCNU, vincristine) | Feb 2016 - Mar 2016 | Craniospinal with posterior fossa boost | 23.4Gy with posterior fossa boost of 54Gy |
| 1496 | Craniospinal Radiotherapy | Medulloblastoma M0 | 4yrs 9m | Completed June 2013 | Standard risk PNET4 guidelines (CCNU, Cisplatin/carboplatin, vincristine) | Apr 2012 - May 2012 | Craniospinal plus posterior fossa boost | 23.4Gy with posterior fossa boost of 30.6Gy |
| 844 | Craniospinal Radiotherapy | Medulloblastoma M0 | 4yrs 1m | Completed Jan 2012 | Standard risk “Packer" chemotherapy (cisplatin, CCNU, vincristine) 6 cycles | Completed Aug 2011 | Craniospinal radiotherapy plus posterior fossa boost | 23.5Gy with posterior fossa boost of 30.6Gy |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1543 | Craniospinal Radiotherapy | Medulloblastoma M3 | 10yrs 1m | T1: 3 cycles Sep 2014 - Nov 2014<br>T2: Jan 2015 - Feb 2015 | T1: St. Jude's high dose medulloblastoma study<br>T2: Cisplatin and Lomustine | Completed Aug 2014 | Craniospinal with posterior fossa boost | 39.6Gy with posterior fossa boost of 16.2Gy |
| 792 | Craniospinal Radiotherapy | Medulloblastoma M3 | 3yrs 6m | T1: Dec 2010 - Apr 2011<br>T2: May 2011 - Jul 2011 | T1: Head Start 3 Protocol<br>T2: High dose chemotherapy and peripheral blood stem cell rescue | July 2011 - Sept 2011 | Craniospinal radiotherapy with posterior fossa boost | 18Gy with conformal boost to posterior fossa of unknown dose |
| 1510 | Craniospinal Radiotherapy | Medulloblastoma M2 | 5yrs 5m | Oct 2015 - Feb 2016 | St Jude's (4 cycles high dose chemotherapy with peripheral blood stem cell rescue) | Jul 2015 - Aug 2015 | Craniospinal tomotherapy with boost to posterior fossa and suprasellar metastasis | 36Gy with posterior fossa and suprasellar metastasis boost of 18Gy |
| 1762 | Craniospinal Radiotherapy | Medulloblastoma M3 | 6yrs 1m | T1: Oct 2015<br>T2: Jan 16 - May 16 | T1: cyclophosphamide<br>T2: CCLG High Risk Medulloblastoma guidelines - St Jude's (SJMB03) - 4 cycles high dose and stem cell rescue | Nov 15 - Dec 15 | Craniospinal with boost to primary site | 39.6Gy with primary site boost to 55.8Gy |
| 1784 | Focal Radiotherapy | Ependymoma | 3yrs 11m | N/A | N/A | July 2017 - Sept 2017 | Radiotherapy | 54Gy |
| 511 | Focal Radiotherapy | Medulloblastoma | 4yrs 1m | Completed Feb 2006 | Treatment according to the UKCCSG Infant PNET Protocol | Completed Aug 2006 | Radiotherapy to posterior fossa | 55Gy to posterior fossa |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 721 | Focal Radiotherapy | Pilocytic Astrocytoma | 3yrs 7m | N/A | N/A | Completed Jul 2013 | Proton beam therapy | Dose unknown |
| 813 | Focal Radiotherapy | Ependymoma | 1yr 5m | Started Mar 2011 | Infant ependymoma protocol, 2 courses given (vincristine, carboplatin then vincristine, methotrexate) | Completed Jul 2011 | Proton beam therapy | Dose unknown |
| 774 | No Radiotherapy | Ependymoma | 2yrs 8m | Sep 10 - Nov 2011 | CCLG infant ependymoma protocol | N/A | N/A | N/A |
| 849 | No Radiotherapy | Medulloblastoma M0 | 1yr 4m | T1: Mar 2011<br>T2: Aug 2011 | T1: Headstart III protocol<br>T2: high dose chemotherapy | N/A | N/A | N/A |
| 594 | No Radiotherapy | Ependymoma | 2yrs 9m | Completed Mar 2009 | CCLG infant ependymoma protocol | N/A | N/A | N/A |
| 1359 | No Radiotherapy | Pilocytic Astrocytoma | 3yrs 3m | N/A | N/A | N/A | N/A | N/A |
| 1672 | No Radiotherapy | Ependymoma | 2yrs 7m | Sep 2004 - Nov 2005 | UKCCSG infant protocol 9204 | N/A | N/A | N/A |
| 860 | No Radiotherapy | Medulloblastoma M2 | 1yr 7m | July 2011 - Dec 2011 | Headstart III protocol | N/A | N/A | N/A |
| 791 | No Radiotherapy | Pilocytic Astrocytoma | 4yrs 6m | N/A | N/A | N/A | N/A | N/A |
| 1459 | No Radiotherapy | Pilocytic Astrocytoma | 8yrs 10m | N/A | N/A | N/A | N/A | N/A |
| 859 | No Radiotherapy | Ependymoma | 0yrs 4m | Sep 2011 - Sep 2012 | Followed CCLG ependymoma guidelines | N/A | N/A | N/A |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1553 | No Radiotherapy | Atypical Choroid Plexus Papilloma | 1yr 7m | N/A | N/A | N/A | N/A | N/A |
| 974 | No Radiotherapy | Pilocytic Astrocytoma | 7yrs 2m | N/A | N/A | N/A | N/A | N/A |
| 1174 | No Radiotherapy | Medulloblastoma M0 | 2yrs 3m | Completed Sep 2013 | Headstart II protocol | N/A | N/A | N/A |

*Table 6.2: A table summarising the treatment received by each of the patients included in the analysis.*

### 6.2.2. Processing Patient Data

DSC-MRI data from the patients was processed in Matlab (The MathWorks, MA, 2019a), using the quality control pipeline developed from earlier work. For further details on how the pipeline functions, and justifications for the techniques used, see the methods section of Chapter 5. As in Chapter 5, four slices were analysed from each dataset, centred around the corpus callosum where possible. Slices higher up in the brain were chosen as these will have received a very low dose during focal radiotherapy.

### 6.2.3. Comparing Treatment Groups

WM and GM ROIs were drawn onto high resolution $T_2$ weighted images and downsampled to the resolution of the DSC images. In the case of GM ROIs, a threshold was applied to remove any CSF from the ROIs. The value of the threshold applied varied between patients. The ROIs were drawn by Stephen Powell using a network created in MeVisLab and were checked by Professor Andrew Peet (a Paediatric Neurooncologist with 18 years of experience) and Dr Stephanie Withey (a Clinical Scientist specialising in MRI with 8 years of experience). From the ROIs it was then possible to calculate average rCBV values from WM and GM. The ratio of average WM rCBV to average GM rCBV ($rCBV_{WM/GM}$) was calculated for each patient and this value was compared between treatment groups. Two sets of comparisons were made: one where the three treatment groups were compared to each other; and one where the focal radiotherapy and no radiotherapy groups were merged into one group and compared to the CS radiotherapy group.

### 6.2.4. Investigating Change in $rCBV_{WM/GM}$ with time

For patients with multiple DSC-MRI scans post-treatment completion, it was possible to plot the change in $rCBV_{WM/GM}$ values with time. Error bars were produced by estimating the standard error in the mean rCBV values from the GM and WM ROIs by dividing the standard deviation in the rCBV by the square root of the number of values within the ROI. The errors

in the average rCBV values from the GM and WM ROIs were then combined to give an error in $rCBV_{WM/GM}$.

### 6.2.5. Statistical Analysis

All statistical testing was carried out using R (R Foundation for Statistical Computing, Vienna, Austria, Version 3.5.0). The Shapiro-Wilk test for normality was used to test all datasets prior to making any comparisons. The Kruskal-Wallis test was used to investigate differences in the SDNR, RMSE, FWHM, PSR and rCBV values before and after denoising, whilst the chi square test was used to investigate difference in the % of signal-time courses passing quality control before and after denoising. Prior to comparing treatment groups, a statistical power analysis was carried out to determine the number of samples required in each group to see a significant difference between the groups. A significance level of 0.05 was selected as this is the significance level used by the statistical tests; an effect size of 0.4 was chosen as this represents a small effect; and a statistical power of 0.8 was used, as this gives a low chance of a type II error occurring (366). ANOVA combined with Tukey post-hoc testing (with a Bonferroni correction) was used to compare the $rCBV_{WM/GM}$ values between treatment groups. A two way ANOVA test was also used to compare the effect of denoising and treatment on patients from the CS and no radiotherapy groups.

## 6.3. Results

### 6.3.1. Effect of Denoising on the Patient Data

For each of the metrics produced by the pipeline, the Shapiro-Wilk test produced a P value < 0.05, showing a non-normal distribution. The denoising step of the quality control pipeline caused an increase in SDNR, a decrease in RMSE and an increase in the % of signal-time courses passing quality control for all datasets. Table 6.3 summarises the % of datasets which saw a significant change post-denoising for each metric. The datasets which did not have a significant improvement in the % of signal-time courses passing quality control post-

denoising, had an average of 55.4 ± 28.6% of signal-time courses passing quality control pre-denoising. Figure 6.1 shows a grouped bar plot showing the average SDNR and % of signal-time courses passing quality control for each treatment group pre- and post-denoising. Two patient datasets were excluded post-denoising as the signal-time courses produced were distorted resulting in a larger rCBV in WM compared to GM. An example GM and WM signal-time course from one of the datasets is shown in Figure 6.2. Both datasets were very low quality pre-denoising with only 14.4% and 12.2% of signal-time courses passing quality control, respectively. For both datasets only 3 components were selected in the time direction by the component selection method used for the Tucker decomposition.

| Metric | % of datasets with a significant change post-denoising |
|---|---|
| SDNR | 92 |
| RMSE | 80 |
| FWHM | 84 |
| PSR | 48 |
| rCBV | 80 |
| % passing quality control | 84 |

*Table 6.3: A summary of the % of datasets that show a significant difference in each metric post-denoising.*

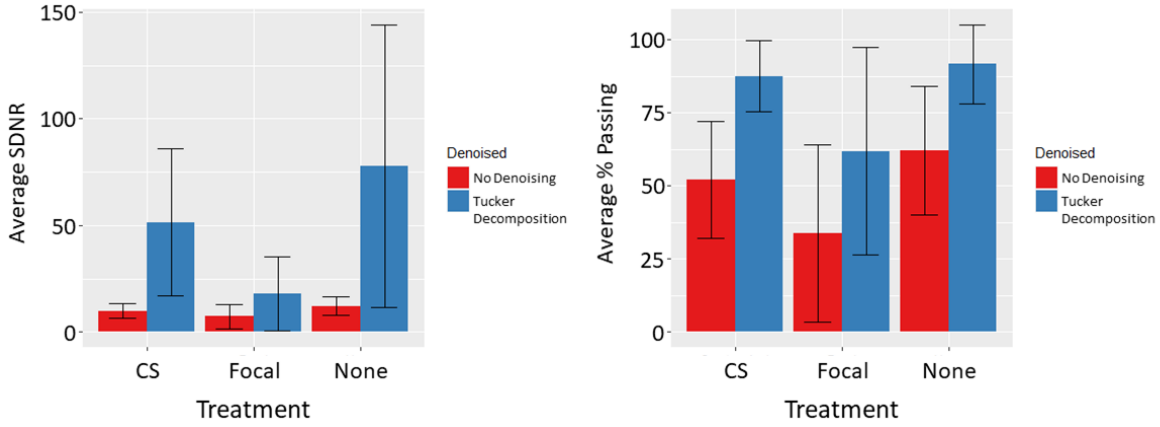



*Figure 6.1: (a) average SDNR for the three treatment groups pre- and post-denoising, (b) average % of signal-time courses passing quality control for the three treatment groups pre- and post-denoising.*

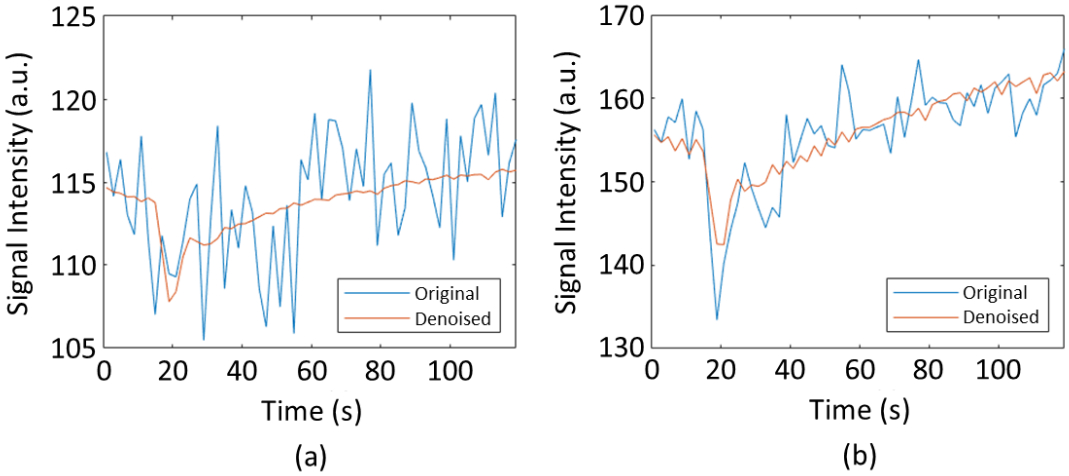



*Figure 6.2: An example (a) WM and (b) GM signal-time course pre- and post-denoising from one of the datasets that was excluded due to issues with the denoising.*

### 6.3.2. Comparing rCBV values between treatment groups

The statistical power analysis showed that for comparing treatments between three groups 21 values would be needed in each group to have 80% power to detect an effect size of 0.4 with $P < 0.05$, whilst for two groups 25 values would be needed to detect an effect size of 0.4 with $P < 0.05$. For each of the treatment groups the Shapiro-Wilk test produced a P value > 0.05, showing a normal distribution of $rCBV_{WM/GM}$ values. Figure 6.3 shows an example of the ROIs, which were used to calculate the average $rCBV_{WM/GM}$. Figure 6.4 shows violin plots of $rCBV_{WM/GM}$ for each of the three treatment groups pre- and post-denoising. Figure 6.5 shows the violin plots for the CS radiotherapy group compared to the combined focal radiotherapy and no radiotherapy groups, pre- and post-denoising.

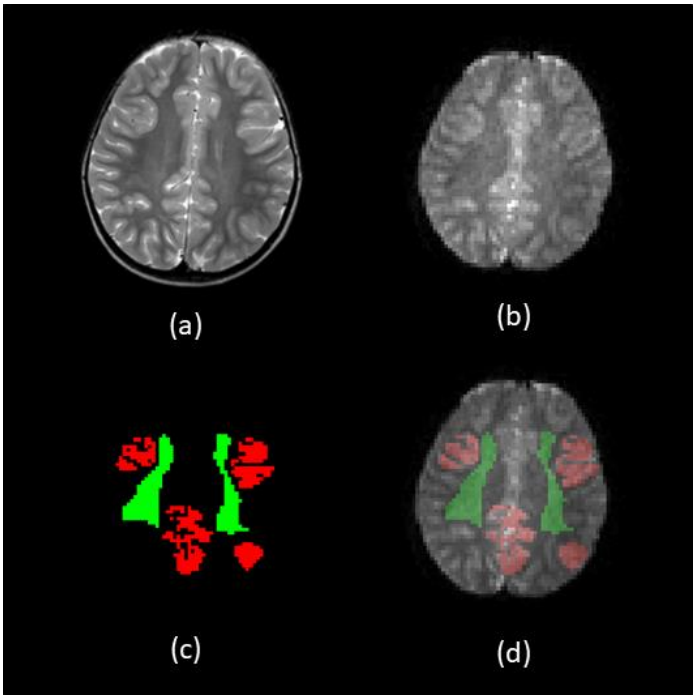



*Figure 6.3: An example of the GM and WM ROIs for one dataset. (a) shows the original $T_2$ image, (b) shows the original DSC image, (c) shows the GM (red) and WM (green) ROIs, and (d) shows the ROIs overlayed onto the DSC.*

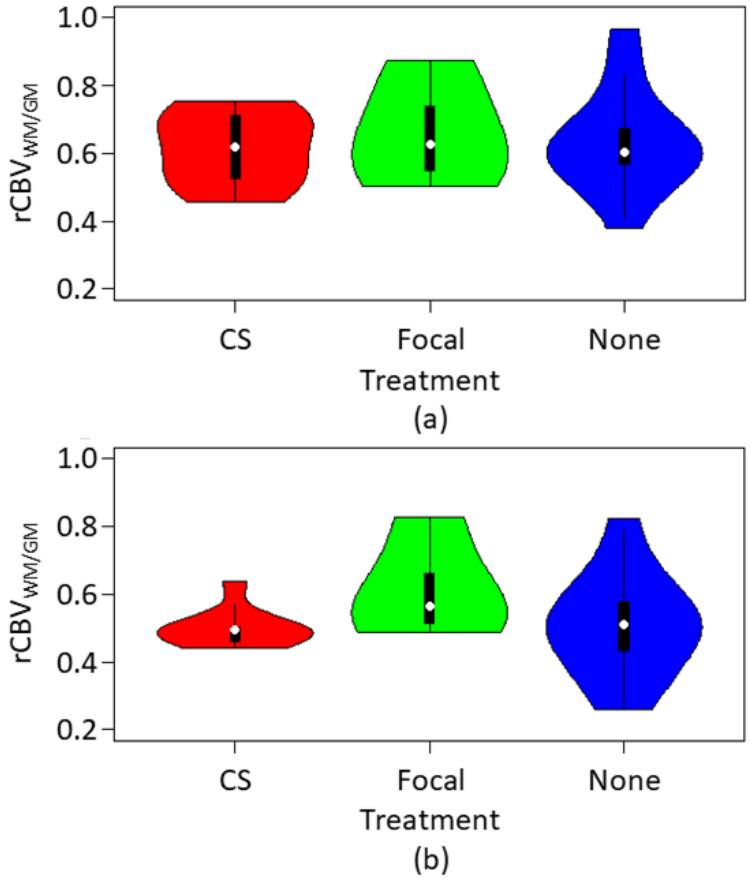



*Figure 6.4: Violin plots showing the differences in the $rCBV_{WM/GM}$ values between the treatment groups (a) before denoising and (b) after denoising.*

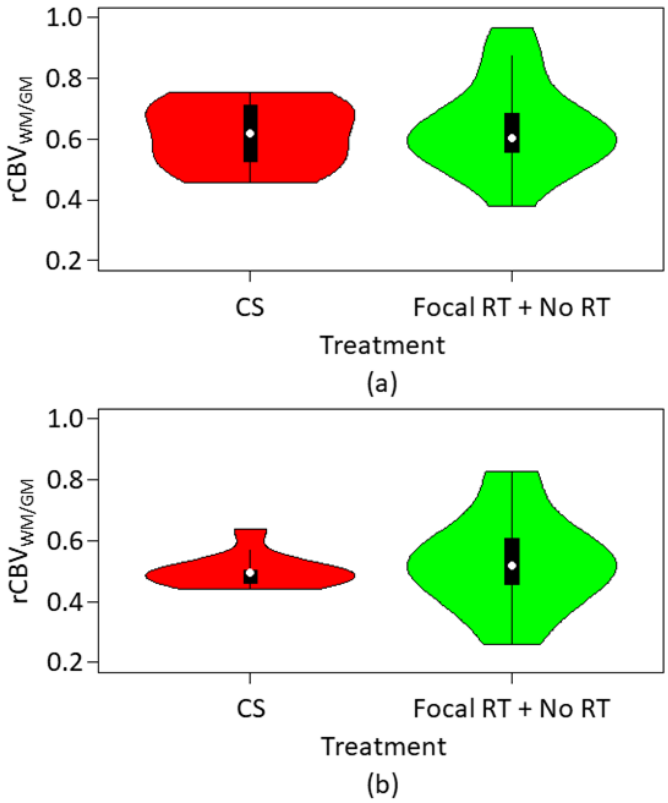



*Figure 6.5: Violin plots showing the differences in $rCBV_{WM/GM}$ between the CS radiotherapy group and the combined focal radiotherapy and no radiotherapy groups, (a) pre-denoising and (b) post-denoising.*

$rCBV_{WM/GM}$ values showed no significant differences between the groups, with ANOVA producing P values of 0.86 and 0.31 pre- and post-denoising for three groups, and P values of 0.64 and 0.63 pre- and post-denoising for two groups. Table 6.4 summarises the P values pre- and post-denoising from running a Tukey post-hoc test on the comparison between three groups.

| Comparison | P Value | |
|---|---|---|
| | Pre-Denoising | Post-Denoising |
| CS-Focal | 0.86 | 0.32 |
| None-Focal | 0.95 | 0.34 |
| None-CS | 0.94 | 0.99 |

*Table 6.4: P values from comparing the effect of different treatment types on $rCBV_{WM/GM}$ values using Tukey post-hoc testing.*

$rCBV_{WM/GM}$ values showed a significant difference due to denoising, but no significant differences between the CS and no radiotherapy groups, with the P values summarised in Table 6.5.

| Category | P Value |
|---|---|
| Treatment | 0.721 |
| Denoising | 0.003 |

*Table 6.5: P values from a two way ANOVA test comparing the effect of treatment and denoising on $rCBV_{WM/GM}$ values.*

Figure 6.6 shows the changes in the ratio of mean $rCBV_{WM/GM}$ for four patients from the CS and focal radiotherapy groups, who were scanned at multiple time points. For all patients there was a decrease in the size of the error bars in $rCBV_{WM/GM}$ except for those from the second scan from CS radiotherapy patient 1. This is because the denoising does not have a significant effect on the data quality, which appears to be due to the data being very poor quality pre-denoising, with only 10.2% of signal-time courses passing quality control.

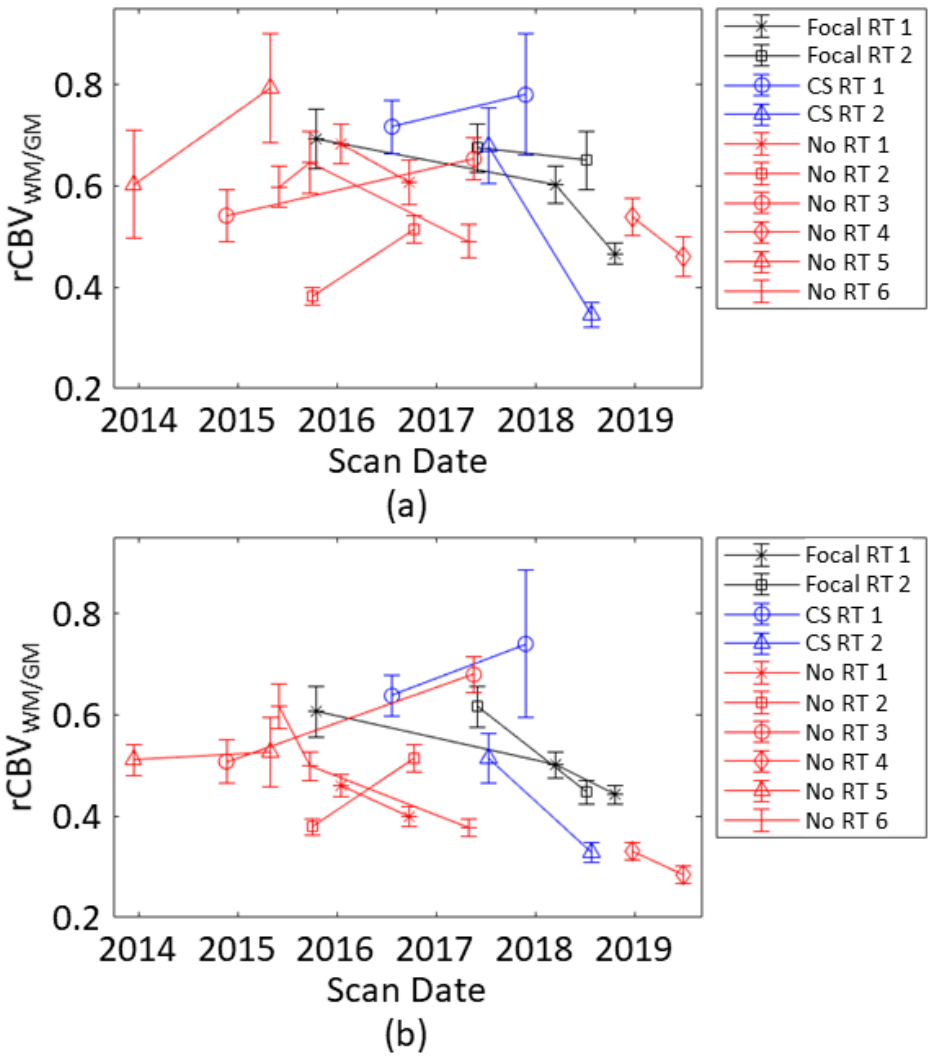



*Figure 6.6: The change in the value of $rCBV_{WM/GM}$ for all patients with scans at multiple time points (a) pre-denoising and (b) post-denoising.*

## 6.4. Discussion

The work in this chapter shows that the processing pipeline developed in this thesis can be applied to patient studies. It also shows that, for this cohort of patients, there were no significant differences in the $rCBV_{WM/GM}$ value in non-tumour brain matter when comparing differing treatment groups.

The denoising step caused a significant change in metric values for the majority of datasets. It led to improved data quality in all datasets as a larger percentage of signal-time courses passed quality control post-denoising. The increase in % passing was statistically significant for 21 out of 25 of the datasets. The average % passing ± the standard deviation shows that there was a large variation in the data quality pre-denoising for the datasets which did not show a significant increase in % passing quality control post-denoising. This varies from the results shown in Chapter 5 and is due to two poor quality datasets not showing a significant improvement in data quality post-denoising. This may be due to the component selection method as this is the key step in Tucker decomposition and in some datasets the method is unable to select the optimal number of components needed to denoise the data.

There is a significant change in the average rCBV post-denoising, but it is not possible to prove whether this change leads to an improved or decreased accuracy in the rCBV, as the 'true' rCBV values in patient data are unknown. However, denoising of simulated data in Chapter 4, where the true value of rCBV is known, showed that the Tucker decomposition produced improved SDNR and improved % error in rCBV values, whilst work in this chapter and Chapter 5 has shown that Tucker decomposition leads to improved SDNR and improved data quality. There is also an expected range of rCBV values for both GM and WM, so it would be possible to compare the expected values to the calculated values to see if they are physiological. If there were a lot of non-physiological values present in a dataset then this

would suggest that there was something wrong with the data. The pipeline does not currently have a method for checking for non-physiological rCBV values.

When the SDNR and % of signal-time courses passing quality control are grouped by treatment type it can be seen that all of the treatment groups have reasonable data quality post-denoising, although the focal radiotherapy group has a smaller percentage of signal-time courses passing quality control than the other groups and a larger standard deviation. The increased range of values is caused by one dataset within the focal radiotherapy group having a larger SDNR and % of signal-time courses passing quality control than the others, due to being acquired with a single-bolus acquisition protocol.

Two datasets were excluded post-denoising due to unusual denoising results, which resulted in $rCBV_{WM/GM}$ values which were greater than 1. GM is more perfused than WM so all $rCBV_{GM/WM}$ values should be less than 1. From the example WM and GM signal-time courses in Figure 6.2 it can be seen that the signal-time courses post-denoising consist of signal-time courses which have an underestimated first pass post-denoising, and signal-time courses that appear to be just noise pre-denoising. As mentioned in the results section the component selection method for the Tucker decomposition only selected three components in the time dimension. This suggests that there is an issue with the current component selection method, which is a key step in Tucker decomposition denoising. Both datasets are very low quality pre-denoising and this may be the reason for the poor performance of the component selection method and therefore the denoising method. Therefore, further work may be needed to develop the component selection method further, to make it more robust to noisy data. Alternatively, it may be necessary to determine a lower quality threshold, below which denoising is not attempted and the data is rejected.

There were no significant differences in $rCBV_{WM/GM}$ values between any of the treatment groups pre- or post-denoising as shown by the ANOVA and Tukey post-hoc testing, for any of the comparisons made. Only denoising had a significant effect on $rCBV_{WM/GM}$ when a two way ANOVA test was applied to the $rCBV_{WM/GM}$ values from the CS and no radiotherapy groups. However, the statistical power analysis shows that more patients are required in each group in order to have sufficient power to detect a significant difference in $rCBV_{WM/GM}$ between groups.

Looking at the violin plots produced, the shape of the distribution changes for each treatment group post-denoising. Each group in the violin plot has a small number of values (4 in the focal radiotherapy group, 9 in the CS radiotherapy group and 12 in the no radiotherapy group), so one or two outlier values would cause a change in the distribution shape. In much larger datasets the outlier values would have less significance and would likely not affect the distribution shape.

The change in $rCBV_{WM/GM}$ values with time for patients with scans acquired at multiple time points shows no overall trend in $rCBV_{WM/GM}$. For the focal radiotherapy group there was an overall decrease in the $rCBV_{WM/GM}$ value for each patient. However, there are only two patients in this group with scans acquired at multiple time points, so this is far from conclusive. For some of the patients the error bars between two scan dates do not overlap, suggesting a larger change in $rCBV_{WM/GM}$ than those where is an overlap. Comparing the plots in Figure 6.6 pre- and post-denoising, it can be seen that denoising does not change any of the trends in $rCBV_{WM/GM}$ between time points, despite causing a significant change in rCBV values. Comparing the error bars pre- and post-denoising it can be seen that denoising causes a decrease in error bar size for all patients and scans, except for the second scan from the first craniospinal radiotherapy patient. This is due to the denoising not having a significant

effect on the data quality, which appears to be due to the data being low quality pre-denoising in this case.

Previous work in this area has shown a difference in perfusion measures between different types of treatment. For example, work by Wenz et al. showed a significant decrease in CBV in normal brain post whole brain radiotherapy and a non-significant decrease post conformal radiotherapy (359). It has also been shown that the significance of the decrease in perfusion measures following focal radiotherapy depends on the dose of radiation received. For example, Fahlstrom et al, showed a dose dependent reduction in CBV and CBF for WM but not for GM (360), whilst Cao et al, found a statistically significant decrease in perfusion measures in regions that had received greater than 20Gy (361), and Fuss et al, found a significant decrease in rCBV in regions of the brain that received more than 40% of the dose (367). The results from this chapter appear to be different from those in the literature. However, this may be due to the methodology. The papers referenced above are all comparing significant differences in perfusion scans on the same patient pre- and post-radiotherapy, whilst this study compares rCBV ratios from groups of patients receiving differing treatments. The DSC-MRI scans in this study are also acquired at differing times following treatment whilst the papers referenced have scans acquired at fixed time periods following treatment, which makes it easier to make comparisons between patients.

In this work the quality control pipeline has been applied to a dataset of paediatric brain tumour patients. It could be applied to assess the data quality and changes in rCBV in DSC-MRI datasets for other types of pathology. The pipeline could also adapted in order to apply it to other perfusion MRI techniques such as dynamic contrast enhanced (DCE-) MRI. This technique is similar to DSC-MRI in that both techniques series involve the acquisition of images are whilst a contrast agent is injected. As the pipeline has essentially been developed

to process signals that vary with time, it could potentially be adapted to be applied to other imaging modalities where a change in signal with time is measured, for example perfusion CT or PET.

There are some limitations to this study. Firstly, this study analyses changes in perfusion with age as a confounding variable. Secondly, the majority of patients did not have any perfusion scans until after treatment began, so there is no way to compare perfusion pre-radiotherapy to post-radiotherapy. Thirdly, there are a limited number of patients in the focal radiotherapy group compared to the other two groups, and the data quality is generally poorer in that group. Ideally, it would be useful to have equally sized groups of high quality data acquired at the same point after treatment, and as seen from the statistical power testing a larger number of patients are needed in each group. Fourthly, the results from the denoising show that the Tucker decomposition component selection method is not fully robust, as it appears to struggle with very poor quality data. Further work is needed to either improve the component selection method or develop a threshold to automatically reject very poor quality data. Finally, there is not currently a method for checking for non-physiological rCBV values, the presence of which would suggest issues with the data or an issue with the automated quality control. Including a method of checking whether rCBV values are physiological would help to further improve the robustness of the pipeline.

## 6.5. Conclusion

This work shows that the quality control pipeline developed in Chapter 5 is applicable to other clinical datasets. It shows that denoising is an important step in ensuring data quality in DSC-MRI data. There were no significant differences in $rCBV_{WM/GM}$ values between the treatment groups tested. However, further work with a larger cohort and longitudinal data

from before treatment is needed to confirm if there are more subtle differences in perfusion following treatment.

# Chapter 7: Conclusions and Future Work

# 7. Conclusions and Future Work

## 7.1. Overview

An automated quality control pipeline for DSC-MRI was developed. The pipeline is capable of segmenting the background and ventricles from a DSC image; improving the data quality through Tucker decomposition denoising; determining the data quality using machine learning applied to metrics extracted from signal-time courses; and generating maps for quality control metrics such as SDNR, RMSE, FWHM, and $T_2$ recovery. It is also capable of producing rCBV maps, with manual input of a WM ROI, which is used to normalise the rCBV values. A simulator capable of modelling DSC-MRI data, with a variety of acquisition protocols was developed and was used to validate the work completed in the denoising chapter.

The developed pipeline has been applied to two differing patient datasets: a multicentre study of paediatric cancer patients to assess variation in data quality with acquisition protocol; and a single-centre study of paediatric cancer survivors who underwent differing types of treatment.

The outputs of this thesis have many potential useful applications. The quality control pipeline is capable of assessing data quality and could be used to apply a more quantitative and automated estimate of data quality when applied to large multicentre datasets. It could be applied to assess data quality and generate perfusion maps in a range of disease states such as age-related diseases or brain trauma. Outside of DSC-MRI, the pipeline could also be adapted to be applied to dynamic contrast enhanced (DCE-) MRI as it is a similar technique where the passage of a contrast agent is imaged, using $T_1$ weighted MRI instead of $T_2^*$ weighted MRI. As the pipeline has been developed to process signal that vary with time, it could potentially be adapted to be applied to other imaging modalities where a change in signal intensity with time is used to calculate estimates of parameters, for example perfusion CT or PET.

However, the pipeline will need some further development to ensure that it is fully robust and fully automated. This may include further developments to the segmentation, denoising and machine learning quality control steps to ensure that they are capable of dealing with a large range of acquisition protocols and disease states and automating the calculation of rCBV values. The DSC-MRI simulator models a variety of differing acquisition protocols and artefacts and will have useful applications in validating any new post-processing methods before they are added to the pipeline.

The quality control pipeline provides a voxel-by-voxel assessment of data quality, but it is important to consider the most effective way to apply the quality assessment results. Excluding individual voxels from a dataset on the basis of their quality could be problematic. This is because the machine learning classifier is not trained on pathology-related voxels, such as those located within brain tumours or areas of low perfusion, and these could be excluded from further analysis. It could also lead to a lot of WM being excluded due to it being less perfused. Therefore, it may be more appropriate to use the quality control pipeline to provide an automated assessment of the overall quality of a dataset, and then use this as a basis to decide whether the entire dataset should be excluded from further analysis or not. Datasets with a large % of signal-time courses passing quality control could be passed as good quality, whilst those with a very low percentage of signal-time courses passing quality control could be excluded as poor quality. For datasets in between these extremes the quality maps could be compared to the standard imaging and the perfusion maps to see if the regions of poor quality matched with expectations.

It is also important to consider the use of denoising and whether it should be applied to every dataset. Currently, denoising is applied to every dataset using the Tucker decomposition. The work produced in Chapters 4, 5, & 6, show that denoising can be used to improve the quality

of both simulated and patient data. Work in Chapter 4 showed that in some cases denoising can cause slight increases in CBV error by missing the bottom of the first pass. However, the overall improvement in data quality outweighs the risks of damaging data. Therefore, denoising should be applied to all datasets which are processed by the quality control pipeline. Some datasets in Chapter 6 did not receive any improvement in quality post-denoising. These datasets were all very poor quality so when using the current version of the quality control pipeline, any poor quality datasets that do not see any improvement in quality should be excluded.

## 7.2. Objectives and Conclusions

This section summarises the main objectives from the introduction chapter, the work completed towards each objective, and the main conclusions drawn from that work.

The first objective was to replace the recommended process for assessing DSC-MRI quality with a new, automated method. This was achieved using a machine learning classifier to assess quality, which was trained using a series of metrics extracted from the signal-time courses and the qualitative review results. The conclusion from this work was that the quality control results produced by the machine learning classifiers were comparable to the difference between the reviewers in the qualitative review process. The significant contribution from this work was that a new, automated method of assessing data quality in DSC-MRI, based on qualitative review, was implemented.

The second objective was to validate the performance of differing denoising methods using simulations. A simulator capable of modelling a variety of acquisition protocols and artefacts was developed and was used to assess matricization, wavelet denoising and Tucker decomposition. The conclusion from this work was that denoising can be used to improve the SDNR in patient and simulated data and reduce the error in CBV in simulated data. However,

frequency based denoising methods appeared to produce distorted signal-time courses. The significant contribution from this work was that a series of denoising methods were validated using a DSC-MRI simulator and then the results were confirmed in patient data.

The third objective was to develop an automated pipeline capable of segmenting the relevant regions of the brain using the DSC-MRI data, denoising the data, assessing data quality and producing estimates of perfusion parameters. An automated segmentation method using features extracted directly from the DSC-MRI signal-time courses was developed, which is capable of segmenting the background and ventricles from both GE-EPI and sPRESTO data. This was combined with the denoising and machine learning work to create an automated pipeline. The conclusion from this work was that the quality control pipeline was capable of automatically assessing quality and that Tucker decomposition provided a significant increase in data quality. The significant contribution from this work was that an automated quality control pipeline was created and applied to a multicentre dataset, where it was used to show the importance of denoising to DSC-MRI data quality, and how a variation in acquisition protocol can affect data quality.

The final objective was to apply the automated pipeline to patient data. The pipeline was applied to two different patient datasets. The first of these was a multicentre dataset consisting of DSC-MRI scans acquired with varying acquisition protocols. This was used to demonstrate that the pipeline can handle large, multicentre datasets with differing acquisition protocols. The conclusion from this dataset was that the acquisition protocol had a significant effect on data quality and that Tucker decomposition is an important step in improving data quality. The second dataset was a single-centre study consisting of paediatric patients with brain tumours who underwent either focal radiotherapy, craniospinal radiotherapy or no radiotherapy. The aim was to determine if the treatment type affected the normalised rCBV in

slices of brain which did not contain tumour. The main conclusions were that: it was possible to apply the quality control pipeline to the datasets and improve the data quality; and that there were no significant differences in rCBV between treatment groups. However, the number of cases was limited particularly in terms of longitudinal data and small changes in perfusion could not be ruled out. The significant contribution from this work was that the quality control pipeline was applied to a new dataset to a clinical dataset and the results suggested that there was no significant difference in rCBV in non-tumour brain between the differing treatment groups.

## 7.3. Limitations and Future Work

There are several limitations to the work produced in this thesis, the majority of which can be addressed by further developing the quality control pipeline. This section will address each of the limitations and discuss potential future work.

The segmentation step of the pipeline functions well on the data it has been tested on so far, and the background segmentation method appears to function well. However, the ventricle segmentation method needs further testing. It has only been compared to the gold standard segmentation method for two datasets. It has not been tested on any slices containing pathology, for example a slice containing brain tumour or a slice containing both brain tumour and ventricle. The automated segmentation method needs to be compared to the gold standard segmentation for a much larger range of data, including slices containing pathology such as brain tumour. This may result in alterations or further development of the segmentation method.

The Tucker decomposition used in the denoising step of the pipeline appears to work well for the majority of the datasets. However, for a minority of datasets the component selection step does not function properly, leading to poor denoising performance. Further investigation into

the component selection method is needed to see if the current method can be replaced by a more robust methods..

When comparing the denoising methods, the wavelet denoising method was applied using assumptions on which type of wavelet, level of denoising, and thresholding method, would be best suited to the data. Further work is needed to ensure that it is fully optimised. The model used to compare the methods could be further developed to produce a more complex model. For example, the model by Mehndiratta et al. includes a term for bolus dispersion (345). Using a more complex model may make it possible to simulate a larger range of artifacts. The model has also only been used to simulate healthy brain. Further developing it to include signal-time course from different types of pathology, could be used to assess the performance of denoising methods in different disease states.

The calculation of metrics step, where SDNR, RMSE, FWHM, PSR and rCBV values are determined functions well and is mostly automated. However, the calculation of rCBV currently requires the manual input of a hand-drawn segmentation for normalisation to WM, which is preventing the pipeline from being completely automated. This could be developed in future work by including an automated segmentation step or by implementing automated AIF detection so that quantitative CBV values could be determined. Furthermore, if the pipeline is going to be applied to disease states such as brain tumours, then it is going to be essential to incorporate leakage correction into the calculation of metrics.

Determination of data quality is currently carried out by a random forest classifier, which has been trained on a dataset of multicentre signal-time courses. Currently, the training dataset only contains 1,027 signal-time courses and does not contain any signal-time courses from brain tumour regions. Therefore, future work could involve creating a new dataset for

training, which would contain a much larger number of signal-time courses along with some signal-time courses from diseased brain. If the new dataset is large enough, it may be possible to investigate replacing the machine learning classifier with deep learning.

The quality control pipeline produces rCBV values but there is currently not a method to check if the values produced are physiological. In patient data the true rCBV value are unknown, so there is not an expected value to compare to. However, there is an expected range of rCBV values for GM and WM so a method for comparing values produced by the pipeline to the expected range could be implemented. This would help to identify non-physiological values and could help improve the robustness of the pipeline.

The work from the final results chapter was centred on a single centre study of brain tumour survivors. The results from this chapter appeared to suggest that there were no significant differences in rCBV between the differing treatment groups. However, there are a number of limitations with the dataset used. Firstly, the number of patients in each treatment group (especially the focal radiotherapy group) were small and the statistical power analysis showed that larger numbers of patients would be needed to see a significant difference between the groups. Secondly, there was variation in the treatment protocols particularly with regards to chemotherapy. Finally, the changes in perfusion were analysed with age as a confounding variable. Therefore, this work could be improved on by repeating the comparison when a larger number of patients have been recruited to the study and there is a more even split across the main treatment groups, and age groups. When investigating the effect of treatment on rCBV values the majority of studies investigate the difference in rCBV at diagnosis and during treatment, whilst this work directly compares rCBV WM to GM ratios between differing treatment groups. Therefore, it may be useful to include patients who have had a

DSC-MRI scan prior to treatment. However, none of the patients in the current study have had DSC-MRI scans prior to treatment starting.

In summary the work carried out in this thesis has led to the development of an automated quality control pipeline and a DSC-MRI simulator. The pipeline has been successfully applied to a multicentre dataset and a study which had an important biomedical question, whilst the simulator has been used to validate denoising work and simulate contrast agent leakage. Both outputs have plenty of potential application in DSC-MRI, with some further work to improve their robustness required.

# List of References


1. Peters AM. The precise physiological definition of tissue perfusion and clearance measured from imaging. Springer; 2018.
2. Thomas DL, Lythgoe MF, Pell GS, Calamante F, Ordidge RJ. The measurement of diffusion and perfusion in biological systems using magnetic resonance imaging. Phys Med Biol. 2000;45(8):R97.
3. Mudaliar AV, Ellis BE, Ricketts PL, Lanz OI, Lee CY, Diller TE, et al. Noninvasive blood perfusion measurements of an isolated rat liver and an anesthetized rat kidney. J Biomech Eng. 2008;130(6).
4. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24(1):27-67.
5. Rief M, Chen MY, Vavere AL, Kendziora B, Miller JM, Bandettini WP, et al. Coronary artery disease: analysis of diagnostic performance of CT perfusion and MR perfusion imaging in comparison with quantitative coronary angiography and SPECT—multicenter prospective trial. Radiology. 2018;286(2):461-70.
6. Grand S, Tahon F, Attye A, Lefournier V, Le Bas J-F, Krainik A. Perfusion imaging in brain disease. Diagn Interv Imaging. 2013;94(12):1241-57.
7. Zoghbi GJ, Dorfman TA, Iskandrian AE. The effects of medications on myocardial perfusion. J Am Coll Cardiol. 2008;52(6):401-16.
8. Stylianopoulos T, Jain RK. Combining two strategies to improve perfusion and drug delivery in solid tumors. Proceedings of the National Academy of Sciences. 2013;110(46):18632-7.
9. Lin W-C, Chen P-C, Huang Y-C, Tsai N-W, Chen H-L, Wang H-C, et al. Dopaminergic therapy modulates cortical perfusion in Parkinson disease with and without dementia according to arterial spin labeled perfusion magnetic resonance imaging. Medicine. 2016;95(5).
10. Drouet C, Morel O, Thines L, Porot C, Boulahdour H. Brain perfusion before and after surgical treatment of moyamoya disease. Clin Nucl Med. 2017;42(5):e269-e72.
11. Cuenod C, Balvay D. Perfusion and vascular permeability: basic concepts and measurement in DCE-CT and DCE-MRI. Diagn Interv Imaging. 2013;94(12):1187-204.
12. Ramalho JN, Fragata IR. 17 Computed Tomography (CT) Perfusion: Basic Principles and Clinical Applications. Vascular Imaging of the Central Nervous System: Physical Principles, Clinical Applications, and Emerging Techniques. 2013;5:257.
13. Allmendinger AM, Tang ER, Lui YW, Spektor V. Imaging of stroke: Part 1, perfusion ct??? Overview of imaging technique, interpretation pearls, and common pitfalls. Am J Roentgenol. 2012;198(1):52-62.
14. Berger A. How does it work?: Positron emission tomography. BMJ: British Medical Journal. 2003;326(7404):1449.

15. Wintermark M, Sesay M, Barbier E, Borbély K, Dillon WP, Eastwood JD, et al. Comparative overview of brain perfusion imaging techniques. Stroke. 2005;36(9):e83-e99.
16. Vaquero JJ, Kinahan P. Positron emission tomography: current challenges and opportunities for technological advances in clinical and preclinical imaging systems. Annu Rev Biomed Eng. 2015;17:385-414.
17. Cosgrove D, Lassau N. Imaging of perfusion using ultrasound. Eur J Nucl Med Mol Imaging. 2010;37(1):65-85.
18. Lee YJ, Kim SH, Kang BJ, Kim YJ. Contrast-enhanced ultrasound for early prediction of response of breast cancer to neoadjuvant chemotherapy. Ultraschall in der Medizin-European Journal of Ultrasound. 2019;40(02):194-204.
19. Porter TR, Xie F. Myocardial perfusion imaging with contrast ultrasound. JACC Cardiovasc Imaging. 2010;3(2):176-87.
20. Pape K, Cusick G, Blackwell RJ, Houang MTW, Sherwood A, Thorburn R, et al. ULTRASOUND DETECTION OF BRAIN DAMAGE IN PRETERM INFANTS. The Lancet. 1979;313(8129):1261-4.
21. Murkin JM, Arango M. Near-infrared spectroscopy as an index of brain and tissue oxygenation. Br J Anaesth. 2009;103(suppl_1):i3-i13.
22. Scheeren T, Schober P, Schwarte L. Monitoring tissue oxygenation by near infrared spectroscopy (NIRS): background and current applications. J Clin Monit Comput. 2012;26(4):279-87.
23. Essig M, Shiroishi MS, Nguyen TB, Saake M, Provenzale JM, Enterline D, et al. Perfusion MRI: the five most frequently asked technical questions. Am J Roentgenol. 2013;200(1):24-34.
24. Jahng G-H, Li K-L, Ostergaard L, Calamante F. Perfusion magnetic resonance imaging: a comprehensive update on principles and techniques. Korean journal of radiology. 2014;15(5):554-77.
25. Morabito R, Alafaci C, Pergolizzi S, Pontoriero A, Bonanno L, Gaeta M, et al. DCE and DSC perfusion MRI diagnostic accuracy in the follow-up of primary and metastatic intra-axial brain tumors treated by radiosurgery with cyberknife. Radiation Oncology. 2019;14(1):1-9.
26. Alsop DC, Detre JA, Golay X, Günther M, Hendrikse J, Hernandez-Garcia L, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn Reson Med. 2015;73(1):102-16.
27. Le Bihan D. What can we see with IVIM MRI? Neuroimage. 2019;187:56-67.
28. Calamante F, Gadian DG, Connelly A. Quantification of perfusion using bolus tracking magnetic resonance imaging in stroke - Assumptions, limitations, and potential implications for clinical use. Stroke. 2002;33(4):1146-51.
29. Calamante F. Perfusion MRI using dynamic-susceptibility contrast MRI: quantification issues in patient studies. Top Magn Reson Imaging. 2010;21(2):75-85.
30. Rohrer M, Bauer H, Mintorovitch J, Requardt M, Weinmann HJ. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. Invest Radiol. 2005;40(11):715-24.

31. Mangla R, Kolar B, Zhu T, Zhong J, Almast J, Ekholm S. Percentage signal recovery derived from MR dynamic susceptibility contrast imaging is useful to differentiate common enhancing malignant lesions of the brain. Am J Neuroradiol. 2011;32(6):1004-10.
32. Cha S, Lu S, Johnson G, Knopp EA. Dynamic susceptibility contrast MR imaging: correlation of signal intensity changes with cerebral blood volume measurements. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2000;11(2):114-9.
33. van Osch MJ. Dynamic susceptibility contrast MRI: acquisition and analysis techniques. Clinical Perfusion MRI: Techniques and Applications. 2013:16.
34. Welker K, Boxerman J, Kalnin A, Kaufmann T, Shiroishi M, Wintermark M. ASFNR recommendations for clinical performance of MR dynamic susceptibility contrast perfusion imaging of the brain. Am J Neuroradiol. 2015;36(6):E41-51.
35. Bleeker EJ, Van Buchem MA, Van Osch MJ. Optimal location for arterial input function measurements near the middle cerebral artery in first-pass perfusion MRI. J Cereb Blood Flow Metab. 2009;29(4):840-52.
36. Bal SS, Yang FPG, Sung Y-F, Chen K, Yin J-H, Peng G-S. Optimal Scaling Approaches for Perfusion MRI with Distorted Arterial Input Function (AIF) in Patients with Ischemic Stroke. Brain sciences. 2022;12(1):77.
37. Jackson A, O'Connor J, Thompson G, Mills S. Magnetic resonance perfusion imaging in neuro-oncology. Cancer Imaging. 2008;8(1):186.
38. Law M, Yang S, Babb JS, Knopp EA, Golfinos JG, Zagzag D, et al. Comparison of cerebral blood volume and vascular permeability from dynamic susceptibility contrast-enhanced perfusion MR imaging with glioma grade. Am J Neuroradiol. 2004;25(5):746-55.
39. Grist JT, Withey S, MacPherson L, Oates A, Powell S, Novak J, et al. Distinguishing between paediatric brain tumour types using multi-parametric magnetic resonance imaging and machine learning: A multi-site study. NeuroImage: Clinical. 2020;25:1-6.
40. Law M, Yang S, Wang H, Babb JS, Johnson G, Cha S, et al. Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging. Am J Neuroradiol. 2003;24(10):1989-98.
41. Maia AC, Malheiros SM, da Rocha AJ, da Silva CJ, Gabbai AA, Ferraz FA, et al. MR cerebral blood volume maps correlated with vascular endothelial growth factor expression and tumor grade in nonenhancing gliomas. Am J Neuroradiol. 2005;26(4):777-83.
42. Knopp EA, Cha S, Johnson G, Mazumdar A, Golfinos JG, Zagzag D, et al. Glial neoplasms: dynamic contrast-enhanced T2*-weighted MR imaging. Radiology. 1999;211(3):791-8.
43. Kickingereder P, Sahm F, Radbruch A, Wick W, Heiland S, Von Deimling A, et al. IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. Sci Rep. 2015;5:16238.

44. Lee S, Choi SH, Ryoo I, Yoon TJ, Kim TM, Lee S-H, et al. Evaluation of the microenvironmental heterogeneity in high-grade gliomas with IDH1/2 gene mutation using histogram analysis of diffusion-weighted imaging and dynamic-susceptibility contrast perfusion imaging. J Neurooncol. 2015;121(1):141-50.
45. Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. Radiology. 2002;222(3):715-21.
46. Cha S, Lupo J, Chen M-H, Lamborn K, McDermott M, Berger M, et al. Differentiation of glioblastoma multiforme and single brain metastasis by peak height and percentage of signal intensity recovery derived from dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. Am J Neuroradiol. 2007;28(6):1078-84.
47. Barajas Jr RF, Phillips JJ, Parvataneni R, Molinaro A, Essock-Burns E, Bourne G, et al. Regional variation in histopathologic features of tumor specimens from treatment-naive glioblastoma correlates with anatomic and physiologic MR Imaging. Neuro Oncol. 2012;14(7):942-54.
48. Lefranc M, Monet P, Desenclos C, Peltier J, Fichten A, Toussaint P, et al. Perfusion MRI as a neurosurgical tool for improved targeting in stereotactic tumor biopsies. Stereotact Funct Neurosurg. 2012;90(4):240-7.
49. Parsons MW, Barber A, Chalk J, Darby DG, Rose S, Desmond P, et al. Diffusion- and perfusion-weighted MRI response to thrombolysis in stroke. Ann Neurol. 2001;51(1):28-37.
50. Butcher KS, Parsons M, MacGregor L, Barber PA, Chalk J, Bladin C, et al. Refining the Perfusion–Diffusion Mismatch Hypothesis. Stroke. 2005;36(6):1153-9.
51. Mattia D, Babiloni F, Romigi A, Cincotti F, Bianchi L, Sperli F, et al. Quantitative EEG and dynamic susceptibility contrast MRI in Alzheimer's disease: a correlative study. Clin Neurophysiol. 2003;114(7):1210-6.
52. Koshimoto Y, Yamada H, Kimura H, Maeda M, Tsuchida C, Kawamura Y, et al. Quantitative analysis of cerebral microvascular hemodynamics with T2-weighted dynamic MR imaging. J Magn Reson Imaging. 1999;9(3):462-7.
53. Fussell D, Young RJ. Role of MRI perfusion in improving the treatment of brain tumors. Imaging Med. 2013;5(5):407–26.
54. Lindner T, Ahmeti H, Juhasz J, Helle M, Jansen O, Synowitz M, et al. A comparison of arterial spin labeling and dynamic susceptibility perfusion imaging for resection control in glioblastoma surgery. Oncotarget. 2018;9(26):18570.
55. Choi JW, Moon W-J. Gadolinium deposition in the brain: current updates. Korean journal of radiology. 2019;20(1):134-47.
56. Roberts DR, Holden KR. Progressive increase of T1 signal intensity in the dentate nucleus and globus pallidus on unenhanced T1-weighted MR images in the pediatric brain exposed to multiple doses of gadolinium contrast. Brain Dev. 2016;38(3):331-6.
57. Miller JH, Hu HH, Pokorney A, Cornejo P, Towbin R. MRI brain signal intensity changes of a child during the course of 35 gadolinium contrast examinations. Pediatrics. 2015;136(6):e1637-e40.

58. Fangusaro J, Witt O, Driever PH, Bag AK, de Blank P, Kadom N, et al. Response assessment in paediatric low-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. The Lancet Oncology. 2020;21(6):e305-e16.
59. Boxerman J, Prah D, Paulson E, Machan J, Bedekar D, Schmainda K. The role of preload and leakage correction in gadolinium-based cerebral blood volume estimation determined by comparison with MION as a criterion standard. Am J Neuroradiol. 2012;33(6):1081-7.
60. Calamante F. Arterial input function in perfusion MRI: a comprehensive review. Prog Nucl Magn Reson Spectrosc. 2013;74:1-32.
61. Zhang J, Liu H, Tong H, Wang S, Yang Y, Liu G, et al. Clinical Applications of Contrast-Enhanced Perfusion MRI Techniques in Gliomas: Recent Advances and Current Challenges. Contrast Media Mol Imaging. 2017;2017:1-27.
62. Khalifa F, Soliman A, El-Baz A, Abou El-Ghar M, El-Diasty T, Gimel'farb G, et al. Models and methods for analyzing DCE-MRI: A review. Med Phys. 2014;41(12):124301.
63. Gordon Y, Partovi S, Müller-Eschner M, Amarteifio E, Bäuerle T, Weber M-A, et al. Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. Cardiovascular diagnosis and therapy. 2014;4(2):147.
64. Lis E, Saha A, Peck KK, Zatcky J, Zelefsky MJ, Yamada Y, et al. Dynamic contrast-enhanced magnetic resonance imaging of osseous spine metastasis before and 1 hour after high-dose image-guided radiation therapy. Neurosurg Focus. 2017;42(1):E9.
65. Donaldson SB, West CM, Davidson SE, Carrington BM, Hutchison G, Jones AP, et al. A comparison of tracer kinetic models for T1-weighted dynamic contrast-enhanced MRI: Application in carcinoma of the cervix. Magn Reson Med. 2010;63(3):691-700.
66. Bisdas S, Naegele T, Ritz R, Dimostheni A, Pfannenberg C, Reimold M, et al. Distinguishing recurrent high-grade gliomas from radiation injury: a pilot study using dynamic contrast-enhanced MR imaging. Acad Radiol. 2011;18(5):575-83.
67. Okuchi S, Rojas-Garcia A, Ulyte A, Lopez I, Ušinskienė J, Lewis M, et al. Diagnostic accuracy of dynamic contrast-enhanced perfusion MRI in stratifying gliomas: A systematic review and meta-analysis. Cancer medicine. 2019;8(12):5564-73.
68. Cao Y, Nagesh V, Hamstra D, Tsien CI, Ross BD, Chenevert TL, et al. The extent and severity of vascular leakage as evidence of tumor aggressiveness in high-grade gliomas. Cancer Res. 2006;66(17):8912-7.
69. Jain R, Narang J, Arbab A, Schultz L, Scarpace L, Mikkelsen T, et al. RA-22. ROLE OF NON-MODEL-BASED SEMI-QUANTITATIVE INDICES OBTAINED FROM DCET1 MR PERFUSION IN DIFFERENTIATING PSEUDO-PROGRESSION FROM TRUE-PROGRESSION. Neuro Oncol. 2011;13(suppl_3).
70. Naish JH, Kershaw LE, Buckley DL, Jackson A, Waterton JC, Parker GJ. Modeling of contrast agent kinetics in the lung using T1-weighted dynamic contrast-

enhanced MRI. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2009;61(6):1507-14.
71. Li X, Welch EB, Chakravarthy AB, Xu L, Arlinghaus LR, Farley J, et al. Statistical comparison of dynamic contrast-enhanced MRI pharmacokinetic models in human breast cancer. Magn Reson Med. 2012;68(1):261-71.
72. Vos EK, Litjens GJ, Kobus T, Hambrock T, Hulsbergen-van de Kaa CA, Barentsz JO, et al. Assessment of prostate cancer aggressiveness using dynamic contrast-enhanced magnetic resonance imaging at 3 T. Eur Urol. 2013;64(3):448-55.
73. de Senneville BD, Mendichovszky IA, Roujol S, Gordon I, Moonen C, Grenier N. Improvement of MRI-functional measurement with automatic movement correction in native and transplanted kidneys. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2008;28(4):970-8.
74. Hackstein N, Heckrodt J, Rau WS. Measurement of single-kidney glomerular filtration rate using a contrast-enhanced dynamic gradient-echo sequence and the Rutland-Patlak plot technique. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2003;18(6):714-25.
75. Kerwin WS, O'Brien KD, Ferguson MS, Polissar N, Hatsukami TS, Yuan C. Inflammation in carotid atherosclerotic plaque: a dynamic contrast-enhanced MR imaging study. Radiology. 2006;241(2):459-68.
76. Zhang N, Zhang L, Qiu B, Meng L, Wang X, Hou BL. Correlation of volume transfer coefficient Ktrans with histopathologic grades of gliomas. J Magn Reson Imaging. 2012;36(2):355-63.
77. Chen J, Yao J, Thomasson D, editors. Automatic determination of arterial input function for dynamic contrast enhanced MRI in tumor assessment. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2008: Springer.
78. Narayanan S, Schmithorst V, Panigrahy A, editors. Arterial Spin Labeling in Pediatric Neuroimaging. Semin Pediatr Neurol; 2020: Elsevier.
79. Havsteen I, Damm Nybing J, Christensen H, Christensen AF. Arterial spin labeling: a technical overview. Acta Radiol. 2018;59(10):1232-8.
80. Sunwoo L, Yun TJ, You S-H, Yoo R-E, Kang KM, Choi SH, et al. Differentiation of glioblastoma from brain metastasis: qualitative and quantitative analysis using arterial spin labeling MR imaging. PLoS One. 2016;11(11):e0166662.
81. Lin L, Xue Y, Duan Q, Sun B, Lin H, Huang X, et al. The role of cerebral blood flow gradient in peritumoral edema for differentiation of glioblastomas from solitary metastatic lesions. Oncotarget. 2016;7(42):69051.
82. Khashbat D, Harada M, Abe T, Ganbold M, Iwamoto S, Uyama N, et al. Diagnostic performance of arterial spin labeling for grading nonenhancing astrocytic tumors. Magn Reson Med Sci. 2018;17(4):277.
83. Kong L, Chen H, Yang Y, Chen L. A meta-analysis of arterial spin labelling perfusion values for the prediction of glioma grade. Clin Radiol. 2017;72(3):255-61.

84. Jovanovic M, Radenkovic S, Stosic-Opincal T, Lavrnic S, Gavrilovic S, Lazovic-Popovic B, et al. Differentiation between progression and pseudoprogresion by arterial spin labeling MRI in patients with glioblastoma multiforme. J BUON. 2017;22:1061-7.
85. Chen Y, Wolk D, Reddin J, Korczykowski M, Martinez P, Musiek E, et al. Voxel-level comparison of arterial spin-labeled perfusion MRI and FDG-PET in Alzheimer disease. Neurology. 2011;77(22):1977-85.
86. Musiek ES, Chen Y, Korczykowski M, Saboury B, Martinez PM, Reddin JS, et al. Direct comparison of fluorodeoxyglucose positron emission tomography and arterial spin labeling magnetic resonance imaging in Alzheimer's disease. Alzheimer's & Dementia. 2012;8(1):51-9.
87. Kamagata K, Motoi Y, Hori M, Suzuki M, Nakanishi A, Shimoji K, et al. Posterior hypoperfusion in Parkinson's disease with and without dementia measured with arterial spin labeling MRI. J Magn Reson Imaging. 2011;33(4):803-7.
88. Razek AAKA, Talaat M, El-Serougy L, Gaballa G, Abdelsalam M. Clinical applications of arterial spin labeling in brain tumors. J Comput Assist Tomogr. 2019;43(4):525-32.
89. Armitage PA, Skipper N, Connolly DJ, Griffiths PD. A qualitative comparison of arterial spin labelling and dynamic susceptibility contrast MRI in 52 children with a range of neurological conditions. The British Journal of Radiology. 2017;90(1069):20160495.
90. Novak J, Withey SB, Lateef S, MacPherson L, Pinkey B, Peet AC. A comparison of pseudo-continuous arterial spin labelling and dynamic susceptibility contrast MRI with and without contrast agent leakage correction in paediatric brain tumours. The British journal of radiology. 2019;92(1094):20170872.
91. Hashido T, Saito S, Ishida T. A radiomics-based comparative study on arterial spin labeling and dynamic susceptibility contrast perfusion-weighted imaging in gliomas. Sci Rep. 2020;10(1):1-10.
92. Suman G, Rusin JA, Lebel RM, Hu HH. Multidelay Arterial Spin Labeling MRI in the Assessment of Cerebral Blood Flow: Preliminary Clinical Experience in Pediatrics. Pediatr Neurol. 2020;103:79-83.
93. Iima M, Le Bihan D. Clinical intravoxel incoherent motion and diffusion MR imaging: past, present, and future. Radiology. 2016;278(1):13-32.
94. Kingsley PB, Monahan WG. Selection of the optimum b factor for diffusion-weighted magnetic resonance imaging assessment of ischemic stroke. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2004;51(5):996-1001.
95. Koh D-M, Collins DJ, Orton MR. Intravoxel incoherent motion in body diffusion-weighted MRI: reality and challenges. Am J Roentgenol. 2011;196(6):1351-61.
96. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology. 1986;161(2):401-7.

97. Federau C, Meuli R, O'Brien K, Maeder P, Hagmann P. Perfusion measurement in brain gliomas with intravoxel incoherent motion MRI. Am J Neuroradiol. 2014;35(2):256-62.
98. Ma W, Zhang G, Ren J, Pan Q, Wen D, Zhong J, et al. Quantitative parameters of intravoxel incoherent motion diffusion weighted imaging (IVIM-DWI): potential application in predicting pathological grades of pancreatic ductal adenocarcinoma. Quantitative imaging in medicine and surgery. 2018;8(3):301.
99. Federau C, Sumer S, Becce F, Maeder P, O'Brien K, Meuli R, et al. Intravoxel incoherent motion perfusion imaging in acute stroke: initial clinical experience. Neuroradiology. 2014;56(8):629-35.
100. Mürtz P, Sprinkart A, Reick M, Pieper C, Schievelkamp A-H, König R, et al. Accurate IVIM model-based liver lesion characterisation can be achieved with only three b-value DWI. Eur Radiol. 2018;28(10):4418-28.
101. Ichikawa S, Motosugi U, Ichikawa T, Sano K, Morisaka H, Araki T. Intravoxel incoherent motion imaging of the kidney: alterations in diffusion and perfusion in patients with renal dysfunction. Magn Reson Imaging. 2013;31(3):414-7.
102. Le Bihan D, Breton E, Lallemand D, Aubin M, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988;168(2):497-505.
103. Paschoal AM, Leoni RF, Dos Santos AC, Paiva FF. Intravoxel incoherent motion MRI in neurological and cerebrovascular diseases. NeuroImage: Clinical. 2018;20:705-14.
104. Meeus EM, Novak J, Withey SB, Zarinabad N, Dehghani H, Peet AC. Evaluation of intravoxel incoherent motion fitting methods in low-perfused tissue. J Magn Reson Imaging. 2017;45(5):1325-34.
105. Zhang S-x, Jia Q-j, Zhang Z-p, Liang C-h, Chen W-b, Qiu Q-h, et al. Intravoxel incoherent motion MRI: emerging applications for nasopharyngeal carcinoma at the primary site. Eur Radiol. 2014;24(8):1998-2004.
106. Yao Y, Zhang S, Tang X, Shi J, Zhu W. Intravoxel incoherent motion diffusion-weighted imaging in stroke patients: initial clinical experience. Clin Radiol. 2016;71(9):938. e11-. e16.
107. Cho GY, Moy L, Zhang JL, Baete S, Lattanzi R, Moccaldi M, et al. Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer. Magn Reson Med. 2015;74(4):1077-85.
108. Radiology ESo. White paper on imaging biomarkers. Insights into imaging. 2010;1(2):42-5.
109. Waterton JC. What imaging biomarkers are and how they are used. Pharmacol Ther. 2001;81:104-7.
110. Dregely I, Prezzi D, Kelly-Morland C, Roccia E, Neji R, Goh V. Imaging biomarkers in oncology: Basics and application to MRI. J Magn Reson Imaging. 2018;48(1):13-26.
111. deSouza NM, Achten E, Alberich-Bayarri A, Bamberg F, Boellaard R, Clément O, et al. Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the EIBALL*

subcommittee of the European Society of Radiology (ESR). Insights into imaging. 2019;10(1):87.
112. Barentsz JO, Weinreb JC, Verma S, Thoeny HC, Tempany CM, Shtern F, et al. Synopsis of the PI-RADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use. Eur Urol. 2016;69(1):41.
113. Zhu YC, Zu DM, Zhang Y, Shan J, Shi XR, Deng SH, et al. A comparative study on superb microvascular imaging and conventional ultrasonography in differentiating BI-RADS 4 breast lesions. Oncol Lett. 2019;18(3):3202-10.
114. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. J Magn Reson Imaging. 2019;49(7):e101-e21.
115. Hoffmann R, von Bardeleben S, ten Cate F, Borges AC, Kasprzak J, Firschke C, et al. Assessment of systolic left ventricular function: a multi-centre comparison of cineventriculography, cardiac magnetic resonance imaging, unenhanced and contrast-enhanced echocardiography. Eur Heart J. 2005;26(6):607-16.
116. Matoba M, Tsuji H, Shimode Y, Nagata H, Tonami H. Diagnostic performance of adaptive 4D volume perfusion CT for detecting metastatic cervical lymph nodes in head and neck squamous cell carcinoma. Am J Roentgenol. 2018:1106-11.
117. Zhang D, Xu A. Application of dual-source CT perfusion imaging and MRI for the diagnosis of primary liver cancer. Oncol Lett. 2017;14(5):5753-8.
118. Hayano K, Shuto K, Koda K, Yanagawa N, Okazumi S, Matsubara H. Quantitative measurement of blood flow using perfusion CT for assessing clinicopathologic features and prognosis in patients with rectal cancer. Dis Colon Rectum. 2009;52(9):1624-9.
119. Win T, Miles KA, Janes SM, Ganeshan B, Shastry M, Endozo R, et al. Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non–small cell lung cancer. Clin Cancer Res. 2013;19(13):3591-9.
120. Nael K, Bauer AH, Hormigo A, Lemole M, Germano IM, Puig J, et al. Multiparametric MRI for differentiation of radiation necrosis from recurrent tumor in patients with treated glioblastoma. Am J Roentgenol. 2018;210(1):18-23.
121. O'connor JP, Jackson A, Parker GJ, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. Nature reviews Clinical oncology. 2012;9(3):167-77.
122. Delgado AF, Delgado AF. Discrimination between glioma grades II and III using dynamic susceptibility perfusion MRI: a meta-analysis. Am J Neuroradiol. 2017;38(7):1348-55.
123. Patel P, Baradaran H, Delgado D, Askin G, Christos P, John Tsiouris A, et al. MR perfusion-weighted imaging in the evaluation of high-grade gliomas after treatment: a systematic review and meta-analysis. Neuro Oncol. 2017;19(1):118-27.
124. van Dijken BR, van Laar PJ, Holtman GA, van der Hoorn A. Diagnostic accuracy of magnetic resonance imaging techniques for treatment response evaluation

in patients with high-grade glioma, a systematic review and meta-analysis. Eur Radiol. 2017;27(10):4129-44.
125. Hara S, Tanaka Y, Ueda Y, Hayashi S, Inaji M, Ishiwata K, et al. Noninvasive evaluation of CBF and perfusion delay of Moyamoya disease using arterial spin-labeling MRI with multiple postlabeling delays: comparison with 15O-gas PET and DSC-MRI. Am J Neuroradiol. 2017;38(4):696-702.
126. Ning Y-C, Ernst RR. Structural identification of organic compounds with spectroscopic techniques: Wiley-VCH Weinheim; 2005.
127. Dale BM, Brown MA, Semelka RC. MRI: basic principles and applications: John Wiley & Sons; 2015.
128. McRobbie DW, Moore EA, Graves MJ, Prince MR. MRI from Picture to Proton: Cambridge university press; 2017.
129. Chen B. Introduction to NMR. 2020.
130. Chrysikopoulos HS. T2 Relaxation (Transverse or Spin–Spin Relaxation). Clinical MR Imaging and Physics: A Tutorial. 2009:11-2.
131. Bernstein MA, King KF, Zhou XJ. Handbook of MRI pulse sequences: Elsevier; 2004.
132. Markl M, Leupold J. Gradient echo imaging. J Magn Reson Imaging. 2012;35(6):1274-89.
133. Nitz W, Reimer P. Contrast mechanisms in MR imaging. Eur Radiol. 1999;9(6):1032-46.
134. Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, techniques, and applications of T2*-based MR imaging and its special applications. Radiographics. 2009;29(5):1433-49.
135. Sijbers J, Scheunders P, Bonnet N, Van Dyck D, Raman E. Quantification and improvement of the signal-to-noise ratio in a magnetic resonance image acquisition procedure. Magn Reson Imaging. 1996;14(10):1157-63.
136. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magn Reson Med. 1995;34(6):910-4.
137. Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2007;26(2):375-85.
138. Association NEM. Determination of signal-to-noise ratio (SNR) in diagnostic magnetic resonance imaging. NEMA Standards Publication MS 1-2001. 2001.
139. Bernstein MA, editor Field Strength Dependence in MRI: Advantages and Artifacts at 3T. Proceedings of the ISMRM; 2006.
140. Runge VM, Nitz WR, Schmeets SH, Clarke G. The physics of clinical MR taught through images. MedPh. 2009;36(5):1925.
141. Graessner J. Bandwidth in MRI. Magnetom Flash. 2013;2:3-8.
142. Froehlich JM. MR contrast agents. How Does MRI Work?: Springer; 2006. p. 103-28.

143. Barbier E. T2-* weighted perfusion MRI. Diagn Interv Imaging. 2013;94(12):1205-9.
144. Sorensen AG, Tievsky AL, Ostergaard L, Weisskoff RM, Rosen BR. Contrast agents in functional MR imaging. J Magn Reson Imaging. 1997;7(1):47-55.
145. Simonsen CZ, Østergaard L, Smith DF, Vestergaard-Poulsen P, Gyldensted C. Comparison of gradient-and spin-echo imaging: CBF, CBV, and MTT measurements by bolus tracking. J Magn Reson Imaging. 2000;12(3):411-6.
146. Pedersen M, Klarhöfer M, Christensen S, Ouallet JC, Østergaard L, Dousset V, et al. Quantitative cerebral perfusion using the PRESTO acquisition scheme. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2004;20(6):930-40.
147. Shiroishi MS, Castellazzi G, Boxerman JL, D'Amore F, Essig M, Nguyen TB, et al. Principles of T2 *-weighted dynamic susceptibility contrast MRI technique in brain tumor imaging. J Magn Reson Imaging. 2015;41(2):296-313.
148. Chavhan GB, Babyn PS, Jankharia BG, Cheng H-LM, Shroff MM. Steady-state MR imaging sequences: physics, classification, and clinical applications. Radiographics. 2008;28(4):1147-60.
149. Lin FH, Huang TY, Chen NK, Wang FN, Stufflebeam SM, Belliveau JW, et al. Functional MRI using regularized parallel imaging acquisition. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2005;54(2):343-53.
150. De Dreu MJ, Schouwenaars I, Rutten G-J, Ramsey NF, Jansma JM. Brain activity associated with expected task difficulty. Front Hum Neurosci. 2019;13:286.
151. Deshmane A, Gulani V, Griswold MA, Seiberlich N. Parallel MR imaging. J Magn Reson Imaging. 2012;36(1):55-72.
152. Gaudino S, Martucci M, Botto A, Ruberto E, Leone E, Infante A, et al. Brain DSC MR perfusion in children: a clinical feasibility study using different technical standards of contrast administration. Am J Neuroradiol. 2019;40(2):359-65.
153. Avula S, Peet A, Morana G, Morgan P, Warmuth-Metz M, Jaspan T. European Society for Paediatric Oncology (SIOPE) MRI guidelines for imaging patients with central nervous system tumours. Childs Nerv Syst. 2021:1-12.
154. Bell LC, Hu LS, Stokes AM, McGee SC, Baxter LC, Quarles CC. Characterizing the influence of preload dosing on percent signal recovery (PSR) and cerebral blood volume (CBV) measurements in a patient population with high-grade glioma using dynamic susceptibility contrast MRI. Tomography. 2017;3(2):89.
155. Kassner A, Annesley D, Zhu X, Li K, Kamaly-Asl I, Watson Y, et al. Abnormalities of the contrast re-circulation phase in cerebral tumors demonstrated using dynamic susceptibility contrast-enhanced imaging: a possible marker of vascular tortuosity. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2000;11(2):103-13.
156. Boxerman J, Schmainda K, Weisskoff R. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. Am J Neuroradiol. 2006;27(4):859-67.

157. Bruening R, Kwong KK, Vevea MJ, Hochberg FH, Cher L, Harsh Gt, et al. Echo-planar MR determination of relative cerebral blood volume in human brain tumors: T1 versus T2 weighting. Am J Neuroradiol. 1996;17(5):831-40.
158. Hu LS, Baxter L, Pinnaduwage D, Paine T, Karis J, Feuerstein B, et al. Optimized preload leakage-correction methods to improve the diagnostic accuracy of dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging in posttreatment gliomas. Am J Neuroradiol. 2010;31(1):40-8.
159. Kaufmann TJ, Smits M, Boxerman J, Huang R, Barboriak DP, Weller M, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. Neuro Oncol. 2020;22(6):757-72.
160. Schmainda K, Prah M, Hu LS, Quarles C, Semmineh N, Rand S, et al. Moving Toward a Consensus DSC-MRI Protocol: Validation of a Low–Flip Angle Single-Dose Option as a Reference Standard for Brain Tumors. Am J Neuroradiol. 2019;40(4):626-33.
161. Schmainda K, Prah M, Wujek R, Connelly J, editors. Moving Towards A DSC-MRI Consensus: A new single dose option for standardized rCBV. Joint Annual Meeting ISMRM-ESMRMB; 2018 16-21 June; Paris, France.
162. Madsen MT. A simplified formulation of the gamma variate function. Phys Med Biol. 1992;37(7):1597.
163. McGehee BE, Pollock JM, Maldjian JA. Brain perfusion imaging: how does it work and what should I use? J Magn Reson Imaging. 2012;36(6):1257-72.
164. Rahimzadeh H, Kazerooni AF, Deevband M, Rad HS. An Efficient Framework for Accurate Arterial Input Selection in DSC-MRI of Glioma Brain Tumors. Journal of biomedical physics & engineering. 2019;9(1):69.
165. Peruzzo D, Bertoldo A, Zanderigo F, Cobelli C. Automatic selection of arterial input function on dynamic contrast-enhanced MR images. Comput Methods Programs Biomed. 2011;104(3):e148-e57.
166. Murase K, Kikuchi K, Miki H, Shimizu T, Ikezoe J. Determination of arterial input function using fuzzy clustering for quantification of cerebral blood flow with dynamic susceptibility contrast-enhanced MR imaging. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2001;13(5):797-806.
167. Chen S, Tyan Y-C, Lai J-J, Chang C-C. Automated determination of arterial input function for dynamic susceptibility contrast MRI from regions around arteries using independent component analysis. Radiol Res Pract. 2016;2016.
168. Leu K, Boxerman JL, Ellingson BM. Effects of MRI Protocol Parameters, Preload Injection Dose, Fractionation Strategies, and Leakage Correction Algorithms on the Fidelity of Dynamic-Susceptibility Contrast MRI Estimates of Relative Cerebral Blood Volume in Gliomas. AJNR Am J Neuroradiol. 2017;38(3):478-84.
169. Kosior RK, Kosior JC, Frayne R. Improved dynamic susceptibility contrast (DSC)-MR perfusion estimates by motion correction. J Magn Reson Imaging. 2007;26(4):1167-72.
170. Vidyasagar R, Abernethy L, Pizer B, Avula S, Parkes LM. Quantitative measurement of blood flow in paediatric brain tumours—a comparative study of

dynamic susceptibility contrast and multi time-point arterial spin labelled MRI. The British journal of radiology. 2016;89(1062):20150624.
171. Dallery F, Bouzerar R, Michel D, Attencourt C, Promelle V, Peltier J, et al. Perfusion magnetic resonance imaging in pediatric brain tumors. Neuroradiology. 2017;59(11):1143-53.
172. Vardal J, Salo RA, Larsson C, Dale AM, Holland D, Groote IR, et al. Correction of B0-Distortions in Echo-Planar-Imaging–Based Perfusion-Weighted MRI. J Magn Reson Imaging. 2014;39(3):722-8.
173. Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from B0 field variations. Magn Reson Med. 1995;34(1):65-73.
174. Chang H, Fitzpatrick JM. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. IEEE Trans Med Imaging. 1992;11(3):319-29.
175. Emblem KE, Bjornerud A, Mouridsen K, Borra RJ, Batchelor TT, Jain RK, et al. T1-and T2*-dominant extravasation correction in DSC-MRI: Part II—predicting patient outcome after a single dose of cediranib in recurrent glioblastoma patients. J Cereb Blood Flow Metab. 2011;31(10):2054-64.
176. Quarles C, Gochberg D, Gore J, Yankeelov T. A theoretical framework to model DSC-MRI data acquired in the presence of contrast agent extravasation. Phys Med Biol. 2009;54(19):5749.
177. Leigh R, Jen SS, Varma DD, Hillis AE, Barker PB. Arrival time correction for dynamic susceptibility contrast MR permeability imaging in stroke patients. PLoS One. 2012;7(12):e52656.
178. Liu HL, Wu YY, Yang WS, Chen CF, Lim KE, Hsu YY. Is Weisskoff model valid for the correction of contrast agent extravasation with combined and effects in dynamic susceptibility contrast MRI? Med Phys. 2011;38(2):802-9.
179. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, 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: An Official Journal of the International Society for Magnetic Resonance in Medicine. 1999;10(3):223-32.
180. Leu K, Boxerman JL, Lai A, Nghiemphu PL, Pope WB, Cloughesy TF, et al. Bidirectional Contrast agent leakage correction of dynamic susceptibility contrast (DSC)-MRI improves cerebral blood volume estimation and survival prediction in recurrent glioblastoma treated with bevacizumab. J Magn Reson Imaging. 2016;44(5):1229-37.
181. Kapitanova K, Son S. Machine learning basics. Intelligent Sensor Networks: CRC Press; 2012. p. 3-29.
182. Chengwang L. Singular Value Decomposition in Active Monitoring Data Analysis. Handbook of Geophysical Exploration: Seismic Exploration. 40: Elsevier; 2010. p. 421-30.
183. Kolda TG, Bader BW. Tensor decompositions and applications. SIAMR. 2009;51(3):455-500.

184. Shehab WA, Al-qudah Z. Singular value decomposition: Principles and applications in multiple input multiple output communication system. Intl J Comput Netwo Commun. 2017;9(1):13-21.
185. Schanze T. Compression and noise reduction of biomedical signals by singular value decomposition. IFAC-PapersOnLine. 2018;51(2):361-6.
186. Guo Q, Zhang C, Zhang Y, Liu H. An efficient SVD-based method for image denoising. IEEE transactions on Circuits and Systems for Video Technology. 2015;26(5):868-80.
187. He Y, Gan T, Chen W, Wang H. Adaptive denoising by singular value decomposition. ISPL. 2011;18(4):215-8.
188. James R, Jolly AM, Anjali C, Michael D, editors. Image denoising using adaptive pca and svd. 2015 Fifth International Conference on Advances in Computing and Communications (ICACC); 2015: IEEE.
189. Lyra-Leite DM, da Costa JPCL, de Carvalho JLA, editors. Improved MRI reconstruction and denoising using SVD-based low-rank approximation. 2012 Workshop on Engineering Applications; 2012: IEEE.
190. Leal N, Zurek E, Leal E. Non-local SVD denoising of MRI based on sparse representations. Sensors. 2020;20(5):1536.
191. Cheng Z, Jing X, He F, Sun G, editors. Denoising differential column image motion lidar signal using singular value decomposition. AOPC 2017: Optical Sensing and Imaging Technology and Applications; 2017: International Society for Optics and Photonics.
192. FAN X-g, WU T-d, ZHI Y-l, Xin W. Denoising Method for Raman Imaging Data Based on Singular Value Decomposition and Median Absolute Deviation. Spectroscopy and Spectral Analysis. 2020;40(2):436.
193. Li Z, Wang L, editors. Denoising of ECG signal during spaceflight using singular value decomposition. International Conference on Space Information Technology 2009; 2010: International Society for Optics and Photonics.
194. Wang Y, Chen H, He D, Peng J, Yuan B, Li Y, editors. Study on improved singular value decomposition de-noising method applied to UAV flight parameter data. 2019 IEEE 20th International Conference on High Performance Switching and Routing (HPSR); 2019: IEEE.
195. Jha SK, Yadava R. Denoising by singular value decomposition and its application to electronic nose data processing. IEEE Sens J. 2010;11(1):35-44.
196. Davies T, Fearn T. Back to basics: the principles of principal component analysis. Spectrosc Eur. 2004;16(6):20.
197. Abdi H, Williams LJ. Principal component analysis. Wiley interdisciplinary reviews: computational statistics. 2010;2(4):433-59.
198. Hernandez W, Mendez A, Göksel T. Application of principal component analysis to image compression. STATISTICS: InTechOpen; 2018. p. 107-37.
199. Frey B. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. 2018.
200. Ledesma RD, Valero-Mora P, Macbeth G. The scree test and the number of factors: a dynamic graphics approach. The Spanish journal of psychology. 2015;18.

201. Gurney-Champion OJ, Collins DJ, Wetscherek A, Rata M, Klaassen R, van Laarhoven HWM, et al. Principal component analysis fosr fast and model-free denoising of multi b-value diffusion-weighted MR images. Phys Med Biol. 2019;64(10):105015.
202. Ulfarsson MO, Solo V. Selecting the number of principal components with SURE. ISPL. 2014;22(2):239-43.
203. Qin SJ, Dunia R. Determining the number of principal components for best reconstruction. Journal of process control. 2000;10(2-3):245-50.
204. Perfilieva I, Hurtik P, editors. Reduction with Application to Pattern Recognition in Large Databases. 2018 IEEE Symposium Series on Computational Intelligence (SSCI); 2018: IEEE.
205. Zhao B, Dong X, Guo Y, Jia X, Huang Y. PCA Dimensionality Reduction Method for Image Classification. Neural Processing Letters. 2022;54(1):347-68.
206. Kaya IE, Pehlivanlı AÇ, Sekizkardeş EG, Ibrikci T. PCA based clustering for brain tumor segmentation of T1w MRI images. Comput Methods Programs Biomed. 2017;140:19-28.
207. Mannfolk P, Wirestam R, Nilsson M, Ståhlberg F, Olsrud J. Dimensionality reduction of fMRI time series data using locally linear embedding. Magn Reson Mater Phys, Biol Med. 2010;23(5):327-38.
208. Malladi SRS, Ram S, Rodríguez JJ. Image denoising using superpixel-based PCA. IEEE Transactions on Multimedia. 2020;23:2297-309.
209. Jade A, Srikanth B, Jayaraman V, Kulkarni B, Jog J, Priya L. Feature extraction and denoising using kernel PCA. Chem Eng Sci. 2003;58(19):4441-8.
210. Kang D, Zhizeng L, editors. A method of denoising multi-channel EEG signals fast based on PCA and DEBSS algorithm. 2012 International Conference on Computer Science and Electronics Engineering; 2012: IEEE.
211. Manjon JV, Coupe P, Buades A. MRI noise estimation and denoising using non-local PCA. Med Image Anal. 2015;22(1):35-47.
212. Kanwal L, Shahid MU, editors. Denoising of 3D magnetic resonance images using non-local PCA and transform-domain filter. 2016 19th International Multi-Topic Conference (INMIC); 2016: IEEE.
213. Zhu H, Zhang J, Wang Z. Arterial spin labeling perfusion MRI signal denoising using robust principal component analysis. J Neurosci Methods. 2018;295:10-9.
214. Balvay D, Kachenoura N, Espinoza S, Thomassin-Naggara I, Fournier LS, Clement O, et al. Signal-to-noise ratio improvement in dynamic contrast-enhanced CT and MR imaging with automated principal component analysis filtering. Radiology. 2011;258(2):435-45.
215. Kolda TG. Multilinear operators for higher-order decompositions. Citeseer; 2006.
216. Brender JR, Kishimoto S, Merkle H, Reed G, Hurd RE, Chen AP, et al. Dynamic Imaging of Glucose and Lactate Metabolism by 13 C-MRS without Hyperpolarization. Sci Rep. 2019;9(1):1-14.
217. Feschet F. Implementation of a denoising algorithm based on high-order singular value decomposition of tensors. Image Processing On Line. 2019;9:158-82.

218. Costa N, Haykin S. Appendix A: An Introduction to Tensor Algebra. 2010.
219. Costa N, Haykin S. Multiple-input multiple-output channel models: theory and practice: John Wiley & Sons; 2010.
220. Vervliet N, Debals O, Sorber L, Van Barel M, De Lathauwer L. Tensorlab 3.0. Mar; 2016.
221. Andersson CA, Bro R. The N-way Toolbox for MATLAB. Chemometrics Intellig Lab Syst. 2000;52(1):1-4.
222. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM journal on Matrix Analysis and Applications. 2000;21(4):1253-78.
223. Andersson CA, Bro R. Improving the speed of multi-way algorithms:: Part I. Tucker3. Chemometrics Intellig Lab Syst. 1998;42(1-2):93-103.
224. Vasilescu MAO, Terzopoulos D, editors. Multilinear image analysis for facial recognition. Object recognition supported by user interaction for service robots; 2002: IEEE.
225. Acar E, Camtepe SA, Krishnamoorthy MS, Yener B, editors. Modeling and multiway analysis of chatroom tensors. International Conference on Intelligence and Security Informatics; 2005: Springer.
226. Muti D, Bourennane S. Multidimensional filtering based on a tensor approach. SigPr. 2005;85(12):2338-53.
227. Bai X, Xu F, Zhou L, Xing Y, Bai L, Zhou J. Nonlocal Similarity Based Nonnegative Tucker Decomposition for Hyperspectral Image Denoising. Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2018;11(3):701-12.
228. Meng SS, Huang LT, Wang WQ. Tensor Decomposition and PCA Jointed Algorithm for Hyperspectral Image Denoising. Ieee Geoscience and Remote Sensing Letters. 2016;13(7):897-901.
229. Yaman B, Weingärtner S, Kargas N, Sidiropoulos ND, Akçakaya M. Low-rank tensor models for improved multidimensional MRI: Application to dynamic cardiac $ t_1 $ mapping. IEEE transactions on computational imaging. 2019;6:194-207.
230. Vargas-Cardona HD, Orozco ÁA, Álvarez AM, Álvarez MA. Tensor decomposition processes for interpolation of diffusion magnetic resonance imaging. Expert Systems with Applications. 2019;118:92-108.
231. Graps A. An introduction to wavelets. IEEE computational science and engineering. 1995;2(2):50-61.
232. Cancino-De-Greiff HF, Ramos-Garcia R, Lorenzo-Ginori JV. Signal de-noising in magnetic resonance spectroscopy using wavelet transforms. Concepts in Magnetic Resonance Part A. 2002;14(6):388-401.
233. Johnstone IM, Silverman BW. Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. The Annals of Statistics. 2004;32(4):1594-649.
234. Singh MK. Denoising of natural images using the wavelet transform: San Jose State University; 2010.
235. Chang SG, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression. ITIP. 2000;9(9):1532-46.

236. Bnou K, Raghay S, Hakim A. A wavelet denoising approach based on unsupervised learning model. EURASIP Journal on Advances in Signal Processing. 2020;2020(1):1-26.
237. Dautov ÇP, Özerdem MS, editors. Wavelet transform and signal denoising using Wavelet method. 2018 26th Signal Processing and Communications Applications Conference (SIU); 2018: Ieee.
238. Tikkanen P. Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal. Biol Cybern. 1999;80(4):259-67.
239. Messer SR, Agzarian J, Abbott D. Optimal wavelet denoising for phonocardiograms. Microelectron J. 2001;32(12):931-41.
240. Ali MN. A wavelet-based method for MRI liver image denoising. Biomedical Engineering / Biomedizinische Technik. 2019;64(6):699.
241. Wink AM, Roerdink JB. Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing. IEEE Trans Med Imaging. 2004;23(3):374-87.
242. Jain AK. Data clustering: 50 years beyond K-means. PaReL. 2010;31(8):651-66.
243. Arthur D, Vassilvitskii S. k-means++: The advantages of careful seeding. Stanford; 2006.
244. Chen J-S, Ching RK, Lin Y-S. An extended study of the K-means algorithm for data clustering and its applications. J Oper Res Soc. 2004;55(9):976-87.
245. Dhanachandra N, Manglem K, Chanu YJ. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science. 2015;54:764-71.
246. Wu M-N, Lin C-C, Chang C-C, editors. Brain tumor detection using color-based k-means clustering segmentation. Third international conference on intelligent information hiding and multimedia signal processing (IIH-MSP 2007); 2007: IEEE.
247. Teichgraeber H, Brandt AR. Systematic comparison of aggregation methods for input data time series aggregation of energy systems optimization problems. Computer Aided Chemical Engineering. 44: Elsevier; 2018. p. 955-60.
248. Jaskowiak PA, Campello RJ, Costa IG, editors. On the selection of appropriate distances for gene expression data clustering. BMC Bioinformatics; 2014: Springer.
249. Oviedo B, Moral S, Puris A. A hierarchical clustering method: Applications to educational data. Intelligent Data Analysis. 2016;20(4):933-51.
250. Govender P, Sivakumar V. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). Atmospheric Pollution Research. 2020;11(1):40-56.
251. Arifin AZ, Asano A. Image segmentation by histogram thresholding using hierarchical cluster analysis. PaReL. 2006;27(13):1515-21.
252. Cordes D, Haughton V, Carew JD, Arfanakis K, Maravilla K. Hierarchical clustering to measure connectivity in fMRI resting-state data. Magn Reson Imaging. 2002;20(4):305-17.
253. Filzmoser P, Baumgartner R, Moser E. A hierarchical clustering method for analyzing functional MR images. Magn Reson Imaging. 1999;17(6):817-26.

254. Vupputuri A, Ashwal S, Tsao B, Ghosh N. Ischemic stroke segmentation in multi-sequence MRI by symmetry determined superpixel based hierarchical clustering. Comput Biol Med. 2020;116:103536.
255. Bishop CM. Neural networks and their applications. Rev Sci Instrum. 1994;65(6):1803-32.
256. Emmert-Streib F, Yang Z, Feng H, Tripathi S, Dehmer M. An introductory review of deep learning for prediction models with big data. Frontiers in Artificial Intelligence. 2020;3:4.
257. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.
258. Bousquet O, von Luxburg U, Rätsch G. Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures: Springer; 2011.
259. Kohavi R, editor A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai; 1995: Montreal, Canada.
260. Perner P. Decision tree induction methods and their application to big data. Modeling and processing for next-generation big-data technologies: Springer; 2015. p. 57-88.
261. Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms: Cambridge university press; 2014.
262. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. Ann Behav Med. 2003;26(3):172-81.
263. Tan P-N, Steinbach M, Kumar V. Classification: basic concepts, decision trees, and model evaluation. Introduction to data mining. 2006;1:145-205.
264. Al-Badarneh A, Najadat H, Alraziqi A. Brain Images Classifier: A Hybrid Approach Using Decision Trees and Genetic Algorithms. JINT. 2016;7(2).
265. Fayaz M, Qureshi MS, Kussainova K, Burkanova B, Aljarbouh A, Qureshi MB. An Improved Brain MRI Classification Methodology Based on Statistical Features and Machine Learning Algorithms. Comput Math Methods Med. 2021;2021.
266. Gholami R, Fakhari N. Support vector machine: principles, parameters, and applications. Handbook of neural computation: Elsevier; 2017. p. 515-35.
267. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction: Springer; 2009.
268. Gupta T, Manocha P, Gandhi TK, Gupta R, Panigrahi BK. Tumor classification and segmentation of MR brain images. arXiv preprint arXiv:171011309. 2017.
269. Sujitha V, Sivagami P, Vijaya M. Support vector machine based epilepsy prediction using textural features of MRI. Procedia Computer Science. 2010;2:283-90.
270. Gonella G, Binaghi E, Nocera P, Mordacchini C, editors. Semi-automatic Segmentation of MRI Brain Metastases Combining Support Vector Machine and Morphological Operators. IJCCI; 2019.
271. Ferreira AJ, Figueiredo MA. Boosting algorithms: A review of methods, theory, and applications. Ensemble machine learning. 2012:35-85.
272. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. The annals of statistics. 2000;28(2):337-407.

273. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A, editors. RUSBoost: Improving classification performance when training data is skewed. 2008 19th international conference on pattern recognition; 2008: IEEE.
274. Minz A, Mahobiya C, editors. MR image classification using adaboost for brain tumor type. 2017 IEEE 7th International Advance Computing Conference (IACC); 2017: IEEE.
275. Savio A, García-Sebastián M, Graña M, Villanúa J, editors. Results of an adaboost approach on Alzheimer's disease detection on MRI. International Work-Conference on the Interplay Between Natural and Artificial Computation; 2009: Springer.
276. Pang H, Liu C, Zhao Z, Wu Y, Wang M, Lin Y, editors. Glioma grading based on gentle-adaboost algorithm and radiomics. 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); 2017: IEEE.
277. Zhang L, Zhao J, Jiang Z, Yang H. Intelligent Measurement of Spinal Curvature Using Cascade Gentle AdaBoost Classifier and Region-Based DRLSE. Journal of Advanced Computational Intelligence and Intelligent Informatics. 2019;23(3):502-11.
278. Zhang B, Zhou H, Wang L, Sung C, editors. Classification based on neuroimaging data by tensor boosting. 2017 International Joint Conference on Neural Networks (IJCNN); 2017: IEEE.
279. Gitto S, Cuocolo R, Annovazzi A, Anelli V, Acquasanta M, Cincotta A, et al. CT radiomics-based machine learning classification of atypical cartilaginous tumours and appendicular chondrosarcomas. EBioMedicine. 2021;68:103407.
280. Maglietta R, Amoroso N, Boccardi M, Bruno S, Chincarini A, Frisoni GB, et al. Automated hippocampal segmentation in 3D MRI using random undersampling with boosting algorithm. Pattern Analysis and Applications. 2016;19(2):579-91.
281. Tong T, Ledig C, Guerrero R, Schuh A, Koikkalainen J, Tolonen A, et al. Five-class differential diagnostics of neurodegenerative diseases using random undersampling boosting. NeuroImage: Clinical. 2017;15:613-24.
282. Breiman L. Random forests. MLear. 2001;45(1):5-32.
283. Payabvash S, Aboian M, Tihan T, Cha S. Machine learning decision tree models for differentiation of posterior fossa tumors using diffusion histogram analysis and structural MRI findings. Front Oncol. 2020;10:71.
284. Yang X, Lei Y, Shu H-K, Rossi P, Mao H, Shim H, et al., editors. Pseudo CT estimation from MRI using patch-based random forest. Medical Imaging 2017: Image Processing; 2017: International Society for Optics and Photonics.
285. Hoffman JIE. Chapter 33 - Logistic Regression. In: Hoffman JIE, editor. Basic Biostatistics for Medical and Biomedical Practitioners (Second Edition): Academic Press; 2019. p. 581-9.
286. Jurafsky D, Martin J. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition2008.

287. McLaren CE, Chen W-P, Nie K, Su M-Y. Prediction of malignant breast lesions from MRI features: a comparison of artificial neural network and logistic regression techniques. Acad Radiol. 2009;16(7):842-51.
288. Xiao R, Cui X, Qiao H, Zheng X, Zhang Y, Zhang C, et al. Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net. Biomed Signal Process Control. 2021;66:102362.
289. Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. Journal of Electronic Science and Technology. 2019;17(1):26-40.
290. Huang Q, Mao J, Liu Y, editors. An improved grid search algorithm of SVR parameters optimization. 2012 IEEE 14th International Conference on Communication Technology; 2012: IEEE.
291. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. Journal of machine learning research. 2012;13(2).
292. Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst. 2012;25.
293. Thornton C, Hutter F, Hoos HH, Leyton-Brown K, editors. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining; 2013.
294. Erickson BJ, Kitamura F. Magician's Corner: 9. Performance Metrics for Machine Learning Models. Radiological Society of North America; 2021. p. e200126.
295. Sebastiani P, Perls TT. Detection of Significant Groups in Hierarchical Clustering by Resampling. Frontiers in Genetics. 2016;7.
296. Song Y-Y, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry. 2015;27(2):130-5.
297. Skurichina M, Duin RP. Bagging, boosting and the random subspace method for linear classifiers. Pattern Analysis & Applications. 2002;5(2):121-35.
298. Subasi A. Chapter 3 - Machine learning techniques. In: Subasi A, editor. Practical Machine Learning for Data Analysis Using Python: Academic Press; 2020. p. 91-202.
299. Cohen J. A Coefficient of Agreement for Nominal Scales Educ Psychol Meas. SAGE Publications Inc. 1960;20:37-46.
300. McHugh ML. Interrater reliability: the kappa statistic. Biochemia medica: Biochemia medica. 2012;22(3):276-82.
301. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth. 2019;22(1):67.
302. Schumacker R, Tomek S. Understanding statistics using R: Springer Science & Business Media; 2013.
303. Hoffman JIE. Chapter 25 - Analysis of Variance. I. One-Way. In: Hoffman JIE, editor. Basic Biostatistics for Medical and Biomedical Practitioners (Second Edition): Academic Press; 2019. p. 391-417.
304. McDonald JH. Handbook of biological statistics: sparky house publishing Baltimore, MD; 2009.

305. Paulson ES, Prah DE, Schmainda KM. Spiral Perfusion Imaging with Consecutive Echoes (SPICE™) for the simultaneous mapping of DSC-and DCE-MRI parameters in brain tumor patients: theory and initial feasibility. Tomography. 2016;2(4):295.
306. Boxerman JL, Rosen BR, Weisskoff RM. Signal-to-noise analysis of cerebral blood volume maps from dynamic NMR imaging studies. J Magn Reson Imaging. 1997;7(3):528-37.
307. Akella NS, Twieg DB, Mikkelsen T, Hochberg FH, Grossman S, Cloud GA, et al. Assessment of brain tumor angiogenesis inhibitors using perfusion magnetic resonance imaging: quality and analysis results of a phase I trial. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2004;20(6):913-22.
308. McKinley R, Hung F, Wiest R, Liebeskind DS, Scalzo F. A machine learning approach to perfusion imaging with dynamic susceptibility contrast MR. Front Neurol. 2018;9:717.
309. Emblem KE, Due-Tonnessen P, Hald JK, Bjornerud A, Pinho MC, Scheie D, et al. Machine learning in preoperative glioma MRI: Survival associations by perfusion-based support vector machine outperforms traditional MRI. J Magn Reson Imaging. 2014;40(1):47-54.
310. Jeong J, Wang L, Ji B, Lei Y, Ali A, Liu T, et al. Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images: Introduction. Quantitative imaging in medicine and surgery. 2019;9(7):1201-13.
311. Ji B, Wang S, Liu Z, Weinberg BD, Yang X, Liu T, et al. Revealing hemodynamic heterogeneity of gliomas based on signal profile features of dynamic susceptibility contrast-enhanced MRI. NeuroImage: Clinical. 2019;23:1-9.
312. Hu LS, Yoon H, Eschbacher JM, Baxter LC, Dueck AC, Nespodzany A, et al. Accurate patient-specific machine learning models of glioblastoma invasion using transfer learning. Am J Neuroradiol. 2019;40(3):418-25.
313. Tsolaki E, Svolos P, Kousi E, Kapsalaki E, Fountas K, Theodorou K, et al. Automated differentiation of glioblastomas from intracranial metastases using 3T MR spectroscopic and perfusion data. Int J Comput Assist Radiol Surg. 2013;8(5):751-61.
314. Swinburne NC, Schefflein J, Sakai Y, Oermann EK, Titano JJ, Chen I, et al. Machine learning for semiautomated classification of glioblastoma, brain metastasis and central nervous system lymphoma using magnetic resonance advanced imaging. Annals of translational medicine. 2019;7(11):232-41.
315. Jung Y. Multiple predicting K-fold cross-validation for model selection. Journal of Nonparametric Statistics. 2018;30(1):197-215.
316. Kane I, Carpenter T, Chappell F, Rivers C, Armitage P, Sandercock P, et al. Comparison of 10 different magnetic resonance perfusion imaging processing methods in acute ischemic stroke: effect on lesion size, proportion of patients with diffusion/perfusion mismatch, clinical scores, and radiologic outcomes. Stroke. 2007;38(12):3158-64.

317. Young GS, Stauss J, Mukundan S. Advanced imaging of adult brain tumors with MRI and PET. Blue Books of Neurology. 36: Elsevier; 2010. p. 71-98.
318. Boxerman JL, Quarles CC, Hu LS, Erickson BJ, Gerstner ER, Smits M, et al. Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in high-grade gliomas. Neuro Oncol. 2020;22(9):1262-75.
319. Thilmann O, Larsson E-M, Björkman-Burtscher I, Ståhlberg F, Wirestam R. Effects of echo time variation on perfusion assessment using dynamic susceptibility contrast MR imaging at 3 tesla. Magn Reson Imaging. 2004;22(7):929-35.
320. Xu Q, Liu Q, Ge H, Ge X, Wu J, Qu J, et al. Tumor recurrence versus treatment effects in glioma: A comparative study of three dimensional pseudo-continuous arterial spin labeling and dynamic susceptibility contrast imaging. Medicine. 2017;96(50).
321. Varallyay CG, Nesbit E, Horvath A, Varallyay P, Fu R, Gahramanov S, et al. Cerebral blood volume mapping with ferumoxytol in dynamic susceptibility contrast perfusion MRI: Comparison to standard of care. J Magn Reson Imaging. 2018;48(2):441-8.
322. Carroll TJ, Rowley HA, Haughton VM. Automatic calculation of the arterial input function for cerebral perfusion imaging with MR imaging. Radiology. 2003;227(2):593-600.
323. Willats L, Calamante F. The 39 steps: evading error and deciphering the secrets for accurate dynamic susceptibility contrast MRI. NMR Biomed. 2013;26(8):913-31.
324. Wirestam R, Stahlberg F. Wavelet-based noise reduction for improved deconvolution of time-series data in dynamic susceptibility-contrast MRI. MAGMA. 2005;18(3):113-8.
325. Murase K, Yamazaki Y, Shinohara M, Kawakami K, Kikuchi K, Miki H, et al. An anisotropic diffusion method for denoising dynamic susceptibility contrast-enhanced magnetic resonance images. Phys Med Biol. 2001;46(10):2713-23.
326. Zhang L, Dong WS, Zhang D, Shi GM. Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recognition. 2010;43(4):1531-49.
327. O'Connor JP, Jackson A, Parker GJ, Jayson GC. DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. Br J Cancer. 2007;96(2):189-95.
328. Murase K, Shinohara M, Yamazaki Y. Accuracy of deconvolution analysis based on singular value decomposition for quantification of cerebral blood flow using dynamic susceptibility contrast-enhanced magnetic resonance imaging. Phys Med Biol. 2001;46(12):3147-59.
329. Vonken EjPA, van Osch MJ, Bakker CJ, Viergever MA. Measurement of cerebral perfusion with dual-echo multi-slice quantitative dynamic susceptibility contrast MRI. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 1999;10(2):109-17.
330. Bjørnerud A, Emblem KE. A fully automated method for quantitative cerebral hemodynamic analysis using DSC–MRI. J Cereb Blood Flow Metab. 2010;30(5):1066-78.

331. Helenius J, Perkio J, Soinne L, Ostergaard L, Carano RA, Salonen O, et al. Cerebral hemodynamics in a healthy population measured by dynamic susceptibility contrast MR imaging. Acta Radiol. 2003;44(5):538-46.
332. Emmer BJ, van Osch MJ, Wu O, Steup-Beekman GM, Steens SC, Huizinga TW, et al. Perfusion MRI in neuro-psychiatric systemic lupus erthemathosus. J Magn Reson Imaging. 2010;32(2):283-8.
333. Grandin CB, Bol A, Smith AM, Michel C, Cosnard G. Absolute CBF and CBV measurements by MRI bolus tracking before and after acetazolamide challenge: Repeatabilily and comparison with PET in humans. Neuroimage. 2005;26(2):525-35.
334. Bjornerud A, Sorensen AG, Mouridsen K, Emblem KE. T1-and T* 2-dominant extravasation correction in DSC-MRI: Part I—theoretical considerations and implications for assessment of tumor hemodynamic properties. J Cereb Blood Flow Metab. 2011;31(10):2041-53.
335. Stanisz GJ, Odrobina EE, Pun J, Escaravage M, Graham SJ, Bronskill MJ, et al. T1, T2 relaxation and magnetization transfer in tissue at 3T. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2005;54(3):507-12.
336. Tofts PS. PD: proton density of tissue water. Quantitative MRI of the Brain: Measuring Changes Caused by Disease Wiley & Sons, Chichester, UK. 2003:85-110.
337. Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. The Journal of molecular diagnostics. 2003;5(2):73-81.
338. Wilcox RR. Introduction to robust estimation and hypothesis testing: Academic press; 2011.
339. Zack GW, Rogers WE, Latt S. Automatic measurement of sister chromatid exchange frequency. J Histochem Cytochem. 1977;25(7):741-53.
340. Misiti M, Misiti Y, Oppenheim G, Poggi J-M. Wavelet toolbox. The MathWorks Inc, Natick, MA. 1996;15:21.
341. Vervliet N, Debals O, De Lathauwer L, editors. Tensorlab 3.0—Numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization. 2016 50th Asilomar Conference on Signals, Systems and Computers; 2016: IEEE.
342. Anandkumar A, Jain P, Shi Y, Niranjan UN, editors. Tensor vs. matrix methods: Robust tensor decomposition under block sparse perturbations. Artificial Intelligence and Statistics; 2016: PMLR.
343. Benou A, Veksler R, Friedman A, Raviv TR. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. Med Image Anal. 2017;42:145-59.
344. Kosior RK, Kosior JC, Frayne R. Improved dynamic susceptibility contrast (DSC)-MR perfusion estimates by motion correction. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2007;26(4):1167-72.
345. Mehndiratta A, Calamante F, MacIntosh BJ, Crane DE, Payne SJ, Chappell MA. Modeling and correction of bolus dispersion effects in dynamic susceptibility contrast MRI. Magn Reson Med. 2014;72(6):1762-74.

346. Knutsson L, Ståhlberg F, Wirestam R. Aspects on the accuracy of cerebral perfusion parameters obtained by dynamic susceptibility contrast MRI: a simulation study. Magn Reson Imaging. 2004;22(6):789-98.
347. Kao YH, Mu-Huo Teng M, Zheng WY, Chang FC, Chen YF. Removal of CSF pixels on brain MR perfusion images using first several images and Otsu's thresholding technique. Magn Reson Med. 2010;64(3):743-8.
348. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. Neuroimage. 2012;62(2):782-90.
349. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging. 2001;20(1):45-57.
350. Dubuisson M-P, Jain AK, editors. A modified Hausdorff distance for object matching. Proceedings of 12th international conference on pattern recognition; 1994: IEEE.
351. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26(3):297-302.
352. Díaz-Pacheco J, Van Delden H, Hewitt R. The Importance of Scale in Land Use Models: Experiments in Data Conversion, Data Resampling, Resolution and Neighborhood Extent. Geomatic approaches for modeling land change scenarios: Springer; 2018. p. 163-86.
353. van Gelderen P, Duyn JH, Ramsey NF, Liu G, Moonen CT. The PRESTO technique for fMRI. Neuroimage. 2012;62(2):676-81.
354. Pollack IF. Brain tumors in children. N Engl J Med. 1994;331(22):1500-7.
355. Aldape K, Brindle KM, Chesler L, Chopra R, Gajjar A, Gilbert MR, et al. Challenges to curing primary brain tumours. Nature Reviews Clinical Oncology. 2019;16(8):509-20.
356. Nguyen QN, Chang EL, Allen PK, Maor MH, Ater JL, Mahajan A, et al. Focal and craniospinal irradiation for patients with intracranial germinoma and patterns of failure. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2006;107(9):2228-36.
357. Pollack IF, Agnihotri S, Broniscer A. Childhood brain tumors: current management, biological insights, and future directions: JNSPG 75th Anniversary Invited Review Article. J Neurosurg Pediatr. 2019;23(3):261-73.
358. Bahl G, White G, Alksne J, Vemuri L, Spear MA. Focal radiation therapy of brain metastases after complete surgical resection. Med Oncol. 2006;23(3):317-24.
359. Wenz F, Rempp K, Hess T, Debus J, Brix G, Engenhart R, et al. Effect of radiation on blood volume in low-grade astrocytomas and normal brain tissue: quantification with dynamic susceptibility contrast MR imaging. AJR American journal of roentgenology. 1996;166(1):187-93.
360. Fahlström M, Blomquist E, Nyholm T, Larsson E-M. Perfusion magnetic resonance imaging changes in normal appearing brain tissue after radiotherapy in glioblastoma patients may confound longitudinal evaluation of treatment response. Radiology and oncology. 2018;52(2):143-51.

361. Cao Y, Tsien CI, Sundgren PC, Nagesh V, Normolle D, Buchtel H, et al. Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for prediction of radiation-induced neurocognitive dysfunction. Clin Cancer Res. 2009;15(5):1747-54.
362. Mankoff DA, Dunnwald LK, Gralow JR, Ellis GK, Schubert EK, Tseng J, et al. Changes in blood flow and metabolism in locally advanced breast cancer treated with neoadjuvant chemotherapy. J Nucl Med. 2003;44(11):1806-14.
363. Hiraga S, Klubes P, Owens ES, Cysyk RL, Blasberg RG. Increases in brain tumor and cerebral blood flow by blood-perfluorochemical emulsion (Fluosol-DA) exchange. Cancer Res. 1987;47(12):3296-302.
364. Wang B, Zhao B, Zhang Y, Ge M, Zhao P, Li C, et al. Absolute CBV for the differentiation of recurrence and radionecrosis of brain metastases after gamma knife radiotherapy: a comparison with relative CBV. Clin Radiol. 2018;73(8):758. e1-. e7.
365. Cao Y, Tsien CI, Nagesh V, Junck L, Ten Haken R, Ross BD, et al. Clinical investigation survival prediction in high-grade gliomas by MRI perfusion before and during early stage of RT. International Journal of Radiation Oncology* Biology* Physics. 2006;64(3):876-85.
366. Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. J Grad Med Educ. 2012;4(3):279.
367. Fuss M, Wenz F, Scholdei R, Essig M, Debus J, Knopp MV, et al. Radiation-induced regional cerebral blood volume (rCBV) changes in normal brain and low-grade astrocytomas: quantification and time and dose-dependent occurrence. International Journal of Radiation Oncology* Biology* Physics. 2000;48(1):53-8.
368. Withey S, MacPherson L, Oates A, Powell S, Novak J, Abernethy L, et al. Multicentre study of perfusion magnetic resonance imaging in paediatric brain tumours. Neuro Oncol. 2019;21(Supplement_4):iv10-iv.

# Appendix 1: K2 Values of Multicentre Dataset

## A1.1. Introduction

The current recommended acquisition and analysis protocols for DSC-MRI of patients with brain tumours includes the use of leakage correction in post-processing (34, 159). The quality control pipeline constructed in Chapter 5 does not include leakage correction. This is because the slices of patient data from the multicentre dataset analysed were selected to avoid tumour. Normal brain has an intact blood brain barrier, so the contrast agent remains within the intravascular space and there is very little contrast agent leakage (59). Therefore, it was felt that contrast agent leakage would not be required. The aim of this appendix is to apply leakage correction to a selection of the datasets used in Chapter 5 and estimate the error which may be introduced by not undertaking leakage correction for these datasets.

## A1.2. Methods

Seven patients were randomly selected from the multicentre dataset and had leakage correction applied to them using the Boxerman method (leakage correction was carried out by Dr Stephanie Withey using in-house software written in the Python programming language (v2.7) (90, 368)), which produced K2 maps for each dataset (156).

A MeVisLab network was created to produce hand-drawn ROIs of 'normal brain'. ROIs were drawn on a high resolution $T_2$ weighted image and then registered down to the DSC dataset. The aim of this was to exclude any artefactual K2 values (caused by blood vessels, CSF, the edge of the brain or the edge of the ventricles). Analysis of the K2 values within the ROIs was carried out using a Matlab script, which calculated the mean and standard deviation in the K2 values within the ROIs for each patient.

One hundred GM signal-time courses were simulated using the DSC-MRI model presented in Chapter 4. Leakage was simulated using the mean K2 values determined from the patient

datasets. rCBV was calculated pre- and post-addition of leakage and this was used to calculate the percentage error in rCBV caused by the contrast agent leakage. ANOVA was used to compare the rCBV values pre- and post-leakage to assess if there was a significant difference. No noise was added as the aim was to investigate the effect of leakage only.

### A1.3. Results

Figure A1.1 shows an example K2 map for one slice from patient 4 alongside the corresponding $T_2$ weighted and DSC images. Figure A1.2 shows the ROIs used to obtain the average and the standard deviation in the K2 values and the K2 map after the ROI has been applied. Table A1.1 summarises the acquisition protocol for each of the patient datasets tested, the mean and standard deviation in K2 and the mean error in K2 for the slices tested. Table A1.2 summarises the results of simulating data with the K2 values from Table A1.1 and shows the mean pre- and post-leakage rCBV values, the P values from comparing the mean rCBV values and the mean percentage error in rCBV caused by the simulated leakage.

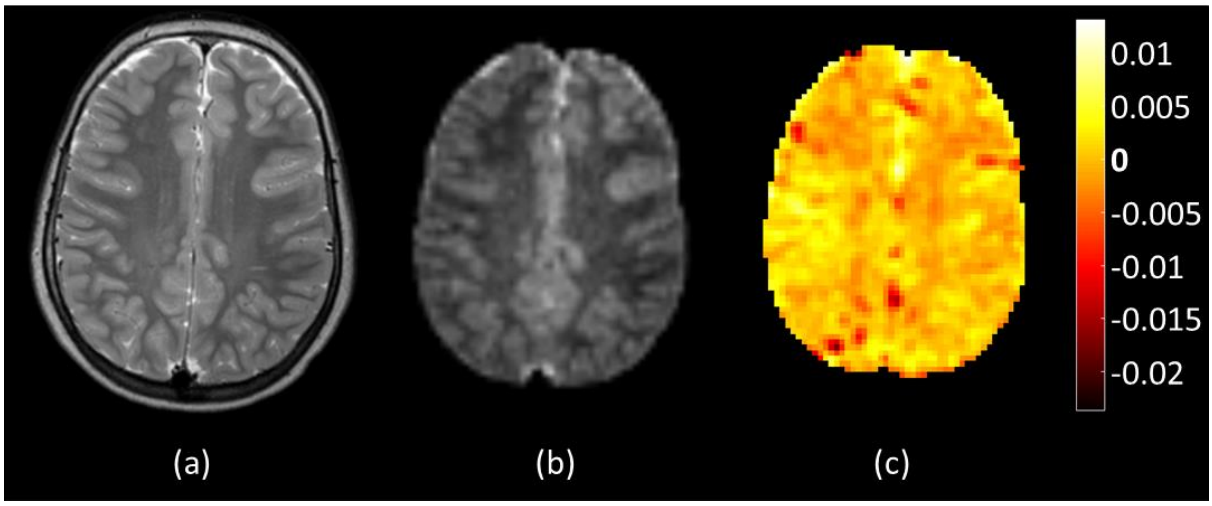



*Figure A1.1: An example of the K2 map produced for one slice of patient 4. (a) shows the $T_2$ weighted image, (b) shows the corresponding slice from the DSC-MRI dataset, and (c) shows the K2 map.*

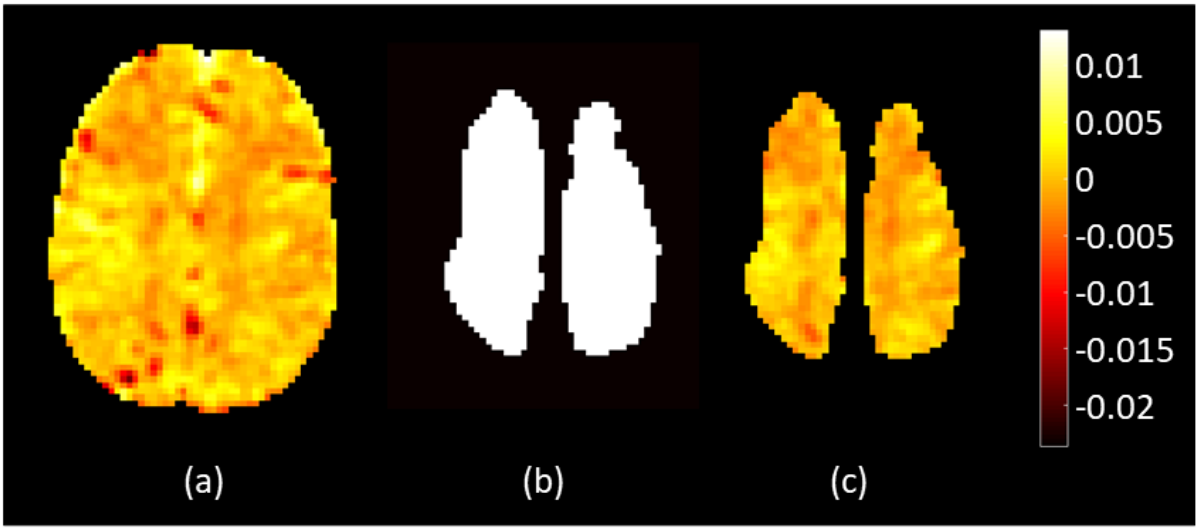



*Figure A1.2 An example of the ROIs drawn and the K2 map following the application of the ROIs from patient 4. (a) shows the original K2 map, (b) shows the ROIs, and (c) shows the K2 map after the application of the ROIs.*

| Patient | Centre | Field Strength | Flip Angle | Protocol | Pre-Bolus | Mean K2 | Std. Dev. K2 | Mean K2 Error |
|---|---|---|---|---|---|---|---|---|
| 1 | Birmingham | 1.5T | 20 | GE-EPI | No | -0.0004 | 0.0027 | 0.0010 |
| 2 | Birmingham | 3T | 7 | sPRESTO | No | -0.0007 | 0.0037 | 0.0012 |
| 3 | Birmingham | 3T | 20 | GE-EPI | Yes | -0.0001 | 0.0039 | 0.0016 |
| 4 | Birmingham | 3T | 20 | GE-EPI | No | -0.0001 | 0.0017 | 0.0006 |
| 5 | Nottingham | 1.5T | 7 | sPRESTO | No | 0.0008 | 0.0040 | 0.0015 |
| 6 | Newcastle | 3T | 45 | GE-EPI | No | 0.0016 | 0.0017 | 0.0008 |
| 7 | Nottingham | 3T | 20 | GE-EPI | Yes | 0.0003 | 0.0028 | 0.0009 |
| 8 | Newcastle | 3T | 45 | GE-EPI | No | 0.0016 | 0.0011 | 0.0005 |

*Table A1.1: A summary of the patient datasets that were tested, their acquisition protocols and the mean and standard deviation in the K2 values.*

| Patient | Mean K2 | Mean rCBV | | | Mean % error in rCBV due to leakage | Std. Dev. in % error in rCBV due to leakage |
|---|---|---|---|---|---|---|
| | | Pre-Leakage | Post-Leakage | P Value from ANOVA | | |
| 1 | -0.0004 | 2.77 | 2.74 | 0.69 | 0.554 | 0.003 |
| 2 | -0.0007 | | 2.75 | 0.46 | 1.055 | 0.003 |
| 3 | -0.0005 | | 2.73 | 0.89 | 0.188 | 0.003 |
| 4 | -0.0001 | | 2.73 | 0.88 | 0.220 | 0.003 |
| 5 | 0.0008 | | 2.69 | 0.35 | 1.316 | 0.004 |
| 6 | 0.0016 | | 2.65 | 0.07 | 2.537 | 0.005 |
| 7 | 0.0017 | | 2.71 | 0.77 | 0.417 | 0.003 |
| 8 | 0.0016 | | 2.65 | 0.07 | 2.525 | 0.005 |

*Table A1.2: A summary of the mean K2 values from the patient datasets and the mean CBV and mean percentage error in CBV from using the K2 values to model leakage in the simulated data.*

## A1.4. Discussion

The work in this appendix suggests that the K2 values from the slices analysed by the quality control pipeline are small. Therefore, not applying leakage correction does not have a significant impact on perfusion measures, as even the largest K2 values only cause a percentage error in rCBV of around 2.5%.

The example K2 map in Figure A1.1 shows that for brain slices which do not contain any brain tumour, the larger K2 values are artificially high in these areas due to their physiology. For these signal-time courses the effects of leakage have been shown to have a negligible effect on the rCBV values.

Hand-drawn ROIs were used to avoid blood vessels, CSF and the edge of the brain or ventricles, as demonstrated in Figure A1.2. This ensures that the analysis of K2 values excludes any artificially high or low values. The hand-drawn ROIs are from slices which do not contain tumour; therefore the K2 values should be zero as no contrast agent leakage is expected. Positive K2 values would suggest $T_1$ weighted contrast agent leakage effects and negative K2 values would suggest $T_2^*$ weighted contrast agent leakage effects (59).

The average K2 values produced from the ROIs are all small and the simulations show that applying any of the K2 values did not cause a significant change in the rCBV values ($P > 0.05$ for all patients), and all of the average % errors in rCBV negligible. However, for patients 6-8 the mean K2 values are a lot larger than the other patients. This difference may be due to differences in the acquisition protocol. For example, patients 6 and 8 were acquired at a centre which used a much larger flip angle than the other centres and did not use a pre-bolus. A larger flip angle results in more $T_1$ weighting so the $T_1$ weighted effects of contrast agent leakage are larger (34). A pre-bolus leaks from the intravascular space to the extravascular extracellular space (EES) to shorten $T_1$ related leakage effects (59). Therefore, for a dataset

with a large flip angle and no pre-bolus, any leakage would result in greater $T_1$ weighted effects on the signal-time courses.

The main limitations of this work are that only a sample of the datasets analysed by the quality control pipeline have been tested by the leakage correction protocol and only one slice from each patient has been analysed. This could be improved by analysing a larger number of datasets. However, if the quality control pipeline is going to be applied to disease states, such as brain tumours, it will be necessary to incorporate leakage correction.

## A1.5. Conclusion

This work confirms that the K2 values in normal brain of the datasets tested is negligible. The simulations carried out show that applying these K2 values results in small changes to the estimated rCBV values. However, if the pipeline is applied to any slices containing brain tumour, then leakage correction should be included in the pipeline.