Arterial Spin Labelling Perfusion MRI: Enhanced Processing and Automated Quality Control

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy, September 2024.

Jian Hu

20289107

Supervised by

Michael Chappell (First Supervisor)

Xin Chen (Second Supervisor)

Signature _____

Date ____/____/____

Abstract

Arterial Spin Labelling (ASL) MRI is the only truly non-invasive method for measuring perfusion using arterial blood water as an endogenous tracer. The settings for ASL MRI acquisitions and analyses have been well developed. However, there are still challenges for conducting research of ASL images.

One challenge is that, due to the relative low resolution of ASL images, partial volume effects could affect the region analysis of ASL images, especially when quantifying perfusion for GM, as the perfusion in voxels could be a confounding of different types of tissues. Accordingly, the approaches to correct partial volume effects have been exploited and applied in prior studies, however, it is absence of the evidence of how partial volume effect corrections could make a difference of analysis in large cohorts.

Another challenge that may limit the widespread clinical adoption of ASL is its reliance on manual Quality Control (QC) by experts. ASL image assessment is traditionally performed by radiologists to ensure data quality for further analysis, a process that is both time-consuming and subjective due to the absence of standardized QC protocols.

In this thesis, for the first challenge, partial volume effects correction was applied in a large ASL data to explore how it affects the region analysis for GM tissues which were intuitively interested in. Apart from the conventional volumetric analysis method, surface-based analysis was also developed and applied to the ASL data to provide in-depth understanding from a different view. Secondly, quality control metrics were developed specific to ASL images, in this thesis, leveraging existing methods and attempting to provide a standardized quality control pipeline for ASL images. Furthermore, deep learning provides the advantages of learning features of medical images without the human interference, provides a promising method to automatically conduct quality control. A deep learning-based method was developed in this thesis for automatic quality control of ASL images.

Acknowledgements

During my four years of study in Nottingham, I learned and experienced a great deal, and completing this thesis would not have been possible without the support of many people. I would like to express my deepest gratitude to all those who helped me along the way.

First, I would like to sincerely thank my primary supervisor, Michael Chappell, for his invaluable guidance and support throughout my studies. His vast knowledge and ability to explain complex concepts in simple terms were instrumental in advancing my research. As the ancient Chinese poet Han Yu once said, “A teacher is one who transmits wisdom, imparts knowledge, and dispels confusion.” Michael exemplified this role, providing not only solutions to immediate challenges but also thoughtful and systematic advice that shaped my approach to research. I feel incredibly fortunate to have had him as my mentor, and his dedication to research will continue to inspire me throughout my career.

I am also deeply grateful to my second supervisor, Xin Chen. Although we worked together for a shorter period, as he became my supervisor in my third year, his impact on my work was profound. His enthusiasm and ability to quickly identify areas for improvement in my research and writing were invaluable. Xin's timely feedback and encouragement helped me develop critical thinking skills and refine my work. I truly appreciate his generous and unwavering support.

I would also like to thank my colleagues and fellow researchers in PhysImAI, Beacon, and SIGMA. Special thanks to Martin Craig, Flora Kennedy McConnell, and Tom Kirk for their technical guidance and for steering my work. I am grateful to my peers, Logan Zhang, Yechuan Zhang, and Haodong He, whose insights and advice from the early stages of my academic journey have been indispensable. Additionally, I want to acknowledge the fellows in SIGMA, whose positivity and support, particularly in the reading group, were a source of inspiration. I am also thankful for the financial support provided by the medicine school and the Haydn Green Foundation.

Finally, I would like to express my heartfelt thanks to my wife Shuya Yang, who took care of me throughout the thesis writing period. Her support has been unwavering, and I am deeply grateful for her presence during this journey.

Contents

List of Tables

List of Figures

Abbreviations

NMR Nuclear Magnetic Resonance

MRI Magnetic Resonance Imaging

ASL Arterial Spin Labelling

CBF Cerebral Blood Flow

GM Gray matter

WM White Matter

CSF Cerebrospinal Fluid

TR Repetition Time

TE Echo Time

SE Spin Echo

GRE Gradient Echo

ATT Arterial Transit Time

aBV arterial Blood Volume

PLD Post-Labelling Delay

QC Quality Control

SVM Support Vector Machine

SNR Signal-to-Noise Ratio

CNR Contrast-to-Noise Ratio

tSNR temporal Signal-to-Noise Ratio

tCNR temporal Contrast-to-Noise Ratio

TI Inversion Time

EPI Echo-planar imaging

cASL continuous Arterial Spin Labelling

pASL pulsed Arterial Spin Labelling

pCASL pseudo-Continuous Arterial Spin Labelling

PVE Partial Volume Effects

PVEc Partial Volume Effects correction

SpCoV Spatial Coefficient of Variation

VAE Variational Autoencoder

GAN Generative Adversarial Network

RICA Right Internal Carotid Artery

LICA Left Internal Carotid Artery

VBA Vertebrobasilar Arteries territory

LS Least Squares

NC Normalized Correlation

MI Mutual Information

NMI Normalized Mutual Information

CR Correlation Ratio

KDE Kernel Density Estimation

MSE Mean Squared Error

CS Cosine Similarity

PLE Poor Labelling Efficiency

DA Delayed Arrival

Mo Motion

LC Low Contrast

PS Poor Signal

1 Introduction

1.1 Background

The brain is a highly dynamic and metabolically active organ that requires a continuous blood supply to maintain its function. This blood supply is referred to as perfusion, a critical physiological process that ensures the delivery of oxygen and nutrients to brain tissue while removing metabolic waste. Arterial Spin Labelling (ASL) MRI is a non-invasive imaging technique to quantitively measure perfusion by the cerebral blood flow (CBF) using magnetically labelled blood water as an endogenous tracer [1] [2] . ASL has been widely applied in neuroscience research to enhance the understanding of brain physiology, assessing cognitive functions, and diagnosing conditions such as stroke, Alzheimer's disease, and other neurodegenerative disorders [2] [3] .

Significant advancements have been performed to translate ASL to clinical practice, including the development of standardized acquisition protocols, image processing and quantification methods that enhance the reliability and reproducibility of ASL across multiple centres with different scanners and sequences. Despite these improvements, ASL suffers from partial volume effects (PVE), where a single voxel may contain a mixture of different tissue types, due to its relatively low spatial resolution [4] . These effects could bias CBF measurements for the brain tissue of interest, especially in aging populations where structural changes in the brain, such as tissue atrophy, are more common. Additionally, ASL is vulnerable to artifacts such as motion, distortion, and poor labelling, which can compromise image quality and lead to inaccuracies in CBF measurements [5] . These issues may hinder the clinical applications of ASL data in routine screening and diagnosis for patients.

Due to the relatively low spatial resolution of ASL, CBF measurements can be significantly affected by PVE, as a single voxel is likely to contain more than one type of tissues. This is particularly important in studies involving patients with tissue atrophy, where changes in tissue composition within voxels might be mistaken for changes in perfusion [6] . Several methods, including the linear regression approach [7] and a spatially regularized Bayesian technique [8] have been developed to correct

PVE in ASL data. However, practical evidence of the impact of PVE correction is still limited, largely due to the small number of studies using these methods and the lack of consistency among them.

Quality control (QC) in ASL MRI is the process of identifying and excluding outliers to ensure that genuine perfusion changes are distinguished from artifacts or low-quality images introduced during acquisition and processing. This step is critical for ensuring the reliability of results before any analysis is undertaken. Traditionally, QC is performed manually, a process that is labor-intensive, subjective, and highly dependent on the expertise of the raters, which is impractical for emerging large-scale datasets [9] . Consequently, there is a growing demand for automated QC tools that can reduce the burden of manual QC, minimize subjectivity, and enhance consistency across large datasets. While many automated pre-processing tools have been developed to improve reliability and reproducibility for ASL MRI, standardized QC protocols remain lacking across these tools. Moreover, QC is not only important during post-processing but also crucial during the ASL acquisition process in clinical settings. If outliers can be detected immediately after acquisition, it allows for the opportunity to re-acquire the scan, ensuring that the data is of high quality from the outset.

Based on the challenges outlined above, this thesis addresses two key issues: first, it aims to fill the gap in the application of PVE correction in a large cohort, exploring how correcting for partial volume effects might accommodate potential changes in grey matter and influence cerebral blood flow measurements. Second, this thesis seeks to standardize QC protocols and develop automated QC tools to detect artifacts in ASL data, thereby facilitating the transition of ASL imaging toward broader clinical use.

1.2 Thesis Outline

This thesis is structured as follows:

Chapter 2 provides the background of this thesis, covering the fundamental principles of MRI, with a focus on the specific artifacts that affect ASL imaging. It details the process of ASL MRI acquisition and pre-processing, highlighting the steps

where quality can be compromised. Furthermore, this chapter reviews existing QC protocols in MRI, with a particular emphasis on metrics relevant to ASL MRI. The potential of machine learning and deep learning techniques used for QC processes are also introduced, setting the stage for the advanced methods developed in later chapters.

Chapter 3 investigates the CBF decline in normal aging by applying partial volume effects correction. Brain atrophy, which could lead to increased PVEs, is a possible factor contributing to the observed decline in CBF, as cortical thinning can result in the contamination of grey matter voxels by other tissue types. This chapter examines how correcting for PVEs, which accounts for changes in grey matter, might affect the understanding of CBF decline in aging populations.

Chapter 4 focuses on developing standardized QC protocols specific to ASL MRI. In this chapter, signal-related metrics, perfusion-related metrics, and registration quality metrics are extracted and evaluated for their effectiveness in identifying artifacts in ASL images. In addition, interactive QC reports were generated using these QC metrics to facilitate the manual screening. Furthermore, machine learning techniques, particularly Support Vector Machines (SVMs), are employed to assess these metrics for identifying outliers, with the goal of developing an automated QC tool for ASL.

Chapter 5 introduces advanced deep learning techniques, specifically a Variational Autoencoder-Generative Adversarial Network (VAE-GAN), which enhances the capabilities of vanilla VAE. This method is employed to detect outliers in ASL MRI data using a deviation-based approach, offering improved performance in identifying anomalies compared to the traditional machine learning method in the former chapter.

Chapter 6 presents the conclusion of the thesis and outlines potential directions for future work.

2 Background

2.1 Magnetic Resonance Imaging

Nuclear Magnetic Resonance (NMR) is a physical phenomenon in which atomic nuclei with magnetic properties respond to an external magnetic field. When placed in a strong static magnetic field, these nuclei align with the field. If a weak oscillating magnetic field is applied, at a frequency matching the nuclei's natural resonance frequency, they absorb and re-emit electromagnetic radiation. This interaction provides detailed information about the local environment of the nuclei, including their chemical surroundings and the strength of the magnetic field they experience. This effect was firstly independently discovered in 1946 by Bloch at Stanford and by Purcell at Harvard [10] , and had since been widely used in chemistry to analyse the composition and structure of molecules.

Over time, advancements in NMR technology led to its adaptation for medical imaging, giving rise to Magnetic Resonance Imaging (MRI). Unlike NMR spectroscopy, which focuses on molecular identification, MRI applies spatial encoding techniques to generate detailed anatomical images of the human body. By exploiting differences in proton density, relaxation times, and tissue composition, MRI enables non-invasive visualisation of soft tissues with exceptional contrast. This transition from NMR to MRI has revolutionised diagnostic imaging, and now it has become more and more prevalent in clinical practice, providing clinicians with a powerful tool for disease detection, diagnosis, and treatment monitoring without ionising radiation.

2.1.1 Principles of MRI

All atomic nuclei are composed of fundamental particles known as protons and neutrons, each possessing an intrinsic property called spin. Atomic nuclei with an even number of both protons and neutrons have a total spin of zero, whereas those with an odd number of either protons or neutrons exhibit a non-zero spin. These nuclei with a non-zero spin have a magnetic moment, which characterizes the

magnetic field surrounding the nucleus with north and south poles, analogous to a bar magnet [10] [11] (see Figure 2.1.1).

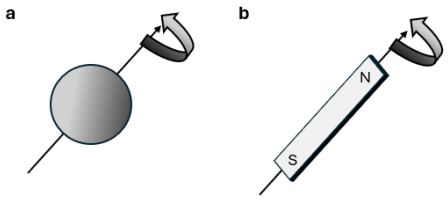


Figure 2.1.1 The magnetic field generated by a nuclear spin.

(a) that is analogous to a bar magnet with north and south poles (b) aligning along its axis of rotation.

NMR can observe the spins of a variety of nuclei (e.g. ^{1}H, ^{13}C, and ^{31}P) depending on the material being studied, while NMR is sensitive to the nuclei's chemical environment [12] . MRI primarily focuses on the spins of hydrogen nuclei (^{1}H) within water molecules to generate contrast, as they are abundant in the tissues of the human body.

In the absence of an external magnetic field, hydrogen nuclei are randomly oriented, leading to no overall orientation when their individual contributions are summed, and thus no detectable signal [13] . However, when a strong external magnetic field (B_0) is applied, hydrogen nuclei experience a torque force, causing their magnetic moments to align parallel to the direction of this field. Consequently, this cause nuclear spins precess about B_0 axis with an angle (see Figure 2.1.2) at a specific frequency, known as the Larmor frequency, which is proportional to the strength of the magnetic field [10] . Notably, when the applied magnetic field is not uniform or there are slight variations in the field across the imaging region, the resonance frequencies of protons may differ slightly, causing off-resonance effects [14] , which may lead to problems in acquisitions, e.g. signal loss or distortion.

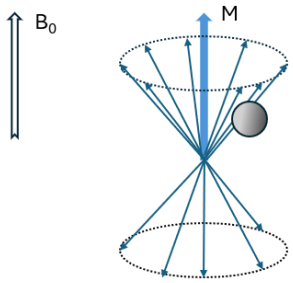


Figure 2.1.2 Nucleus precessing around an external magnetic field (B_0).

After the external magnetic field (B_0) applied, magnetic moments of hydrogen nuclei were aligned, parallel to the direction of this field. The transverse component remains zero due to the random phase of the magnetic moments.

The longitudinal magnetic moment of spins can only be either parallel (spin-up) or antiparallel (spin-down) with B_0, corresponding to a low-energy state and a high-energy state, respectively. A dynamic balance between the magnetic field and temperature determines the two basic energy states. Spins in the low-energy state can absorb energy from an external source and transition to the high-energy state, while spins in the high-energy state can release the same amount of energy and return to the low-energy state [13] . This process is also known as radiofrequency (RF) excitation. Nuclei reach thermal equilibrium when the number of transitions between the lower and upper states is equal in both directions. The resulting magnetization at this point is known as equilibrium magnetization (M_0).

MR signals can only be detected when transverse magnetisation, oriented perpendicular to B_0, is produced. This is achieved by applying another RF field (B_1) rotating in sync with the precessing spins (perpendicular to B_0) to tip the spins away from the longitudinal axis and into the transverse plane. The degree of this tip, or flip angle, is determined by the duration and amplitude of the RF pulse: a 90° RF pulse rotates the entire net magnetisation into the transverse plane, while a 180° RF pulse inverts it to the opposite direction along the longitudinal axis [13] .

The MR signal is generated by the induced voltage change caused by the refocused transverse magnetization, known as the echo. An MR pulse sequence comprises a series of RF pulses and gradients that are precisely timed to produce the desired echo. After an initial 90° RF pulse is applied, the longitudinal spins are tipped into

the transverse plane, as shown in Figure 2.1.3. This induces an oscillating voltage in a receiver coil according to Faraday's law of induction [10] , producing the free-induction decay (FID) signal. Following this, if a 180° pulse is applied after a certain time duration, the phase of the spins is reversed, refocusing the transverse magnetization after the same time duration. The resulting signal is known as the spin echo (SE) [15] , while another commonly used pulse sequence, the gradient-echo (GRE) [16] , employs a refocusing gradient RF pulse instead of a 90° RF pulse. The time between the excitation pulse and the peak of the echo is the echo time (TE), while the time between consecutive excitation pulses is the repetition time (TR). TE and TR are key parameters in an MR pulse sequence, as they govern the extent of relaxation in the transverse and longitudinal directions, respectively. In contrast, the longitudinal magnetization in thermal equilibrium remains static and does not generate a signal.

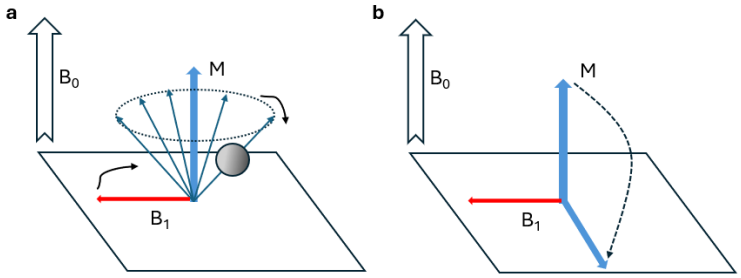


Figure 2.1.3 The generation of the transverse magnetization.

Longitudinal magnetization is tipped into the transverse plane by a B_1 magnetic field from an RF pulse. For effective tipping, the B_1 field must rotate in sync with the precessing transverse magnetization, achieved by a circularly polarized RF pulse at the Larmor frequency. A sufficiently strong and prolonged RF pulse can tip all the longitudinal magnetization by 90 degrees into the transverse plane.

Relaxation refers to the process by which nuclear spins return to thermal equilibrium after absorbing RF energy. There are two types: longitudinal and transverse relaxation, characterized by the time constants T1 and T2 [17] , respectively.

During RF stimulation, nuclei absorb energy and move to an excited state, returning to the ground state by dissipating this energy to their surroundings, known as the lattice. T1 relaxation, or spin-lattice relaxation, describes the recovery of longitudinal magnetization toward equilibrium. The T1 time varies depending on the main

magnetic field strength according to the Larmor frequency, representing the time it takes for transverse magnetization to recover to 63% of its original value [13] . In the meantime, transverse magnetization decays as the magnetic moments lose phase coherence due to mutual interactions. This is because the hydrogen nuclei could be in slightly different local environments (e.g. based on the proximity to other molecules), and they could have different resonant frequencies, leading to the transverse net magnetization disappearing over time, as the faster ones start to cancel out the slower ones, which is known as dephasing. Any changes in magnetic field strength (e.g. B_0 inhomogeneity) affect the frequency of precession could cause dephasing and a reduction in transverse magnetization. This process, known as T2 relaxation, which varies across different tissues, indicates how quickly transverse magnetization diminishes, and corresponds to the time it takes for the transverse magnetization to retain 37% of its original value [17] . Dephasing of spins caused by B_0 field inhomogeneities may lead to a faster $T2^*$ decay when using GRE, as smaller flip angle (<90°) leads to some retention of longitudinal magnetisation. As a result, the build-up time for longitudinal magnetisation is reduced. The difference between $T2^*$ and T2 can be compensated for using SE that is applied halfway of TE to flip the transverse net magnetization by 180°, and the phase of tipped spins will be refocused at TE. T2 decay is generally 5-10 times slower than T1 decay [31] . Since T1 and T2 relaxation times are tissue-specific (as illustrated in Figure 2.1.4), varying the combinations of TE and TR allows for the manipulation of contrast in T1-weighted, T2-weighted, and proton-density (PD) weighted MR images (Figure 2.1.5).

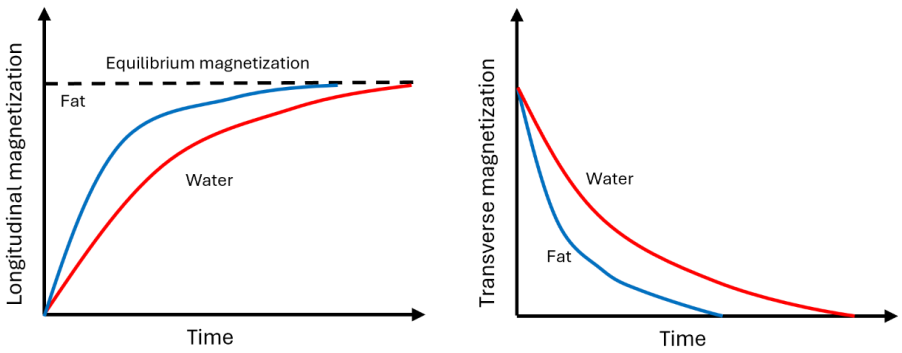


Figure 2.1.4 T_1 and T_2 relaxations in different types of tissues.

Left: T_1 recovery in fat (blue line) and water (red line) illustrates that a short T_1 value, as seen in fat, results in rapid restoration of longitudinal magnetization, while a longer T_1 value, as in water,

indicates slower recovery of magnetization. Right: T2 relaxation in fat (blue line) and water (red line). A shorter T2 value indicates that transverse magnetization decays quickly, while a longer T2 value signifies slower recovery of the magnetization.

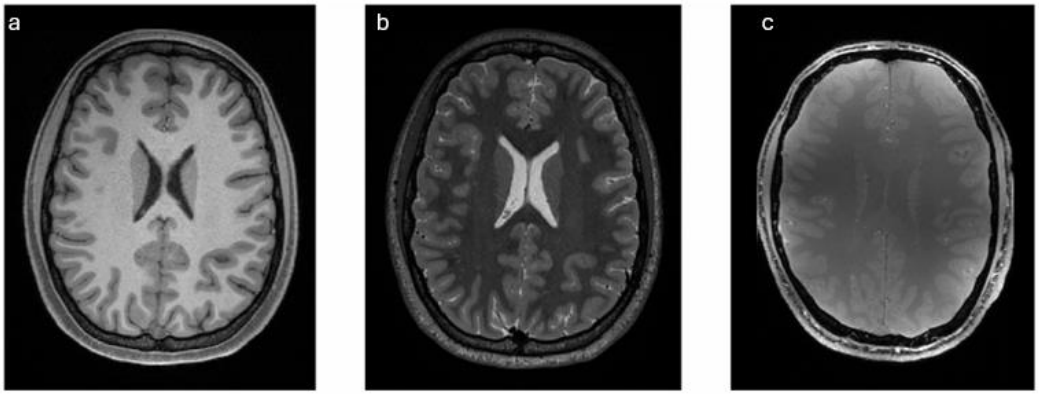


Figure 2.1.5 Different settings of T_1 and T_2 provides different contrast images from [31] .

(a) Short T1 and short T2 generate T1-weighted image. (b) Long T1 and long T2 generate T2-weighted image. (c) Long T1 and short T2 generate PD-weighted image, where the MR signal depends predominantly on the density of the protons, instead of T1 and T2 contrast between tissues.

As mentioned earlier, the resonance frequency of nuclear spins is linked to the strength of the external magnetic field. To localize the MR signal within different areas of tissue, magnetic field gradients are required. These gradients create spatially linear variations in the static field strength and can be applied in any of the three orthogonal directions using the gradient coils in the MR system. Variations in precession speed are detected as higher or lower MR signals, allowing frequency measurements to differentiate MR signals from different spatial locations and enabling 3D image reconstruction. This process is known as spatial encoding [13] .

The first step in spatial encoding is slice selection. By applying a gradient along the main magnetic field (B_0), the magnetic field strength will vary linearly along that direction. When a specific RF pulse is applied, only the nuclei in a slice where the magnetic field strength matches the frequency of the RF pulse will resonate and produce a signal. This way, the MRI system can select a specific slice of the body to image. After selecting a slice, the MRI needs to encode information within that slice. This is done through frequency encoding: another gradient (known as readout gradient) applied by a 180° refocusing pulse along, for example, x-axis within the slice during the acquisition of the MR signal. This causes the precession frequency of the hydrogen nuclei to vary linearly across the slice. As a result, nuclei at different

locations along x-axis resonate at different frequencies. To encode the second spatial dimension within the slice (i.e., the y-axis), the MRI uses phase encoding. This process involves applying a gradient along the y-axis for a short duration before signal acquisition. Phase encoding consists of multiple steps, each introducing a phase shift between adjacent spins. At the edges of the phase encoding gradient, spins undergo a 180° phase shift from the previous step, while those near the isocenter experience a smaller phase shift. This temporarily modifies the phase of spinning nuclei at different y-axis locations. Once the gradient is turned off, all nuclei resume precessing at the same frequency, but their phase differences remain, depending on their y-coordinates.

The signals collected by the MRI, which now contain spatial information encoded through these gradients, are stored in a data matrix called k-space [30] (see Figure 2.1.7). The number of rows in k-space corresponds to the number of phase encoding steps, while the number of columns represents the sampling points along the frequency encoding direction. Thus, the total scan time is proportional to the TR and the number of phase encoding steps used. Data with lower phase encoding gradient amplitudes populate the central rows of k-space, which primarily contain information about the overall structure and contrast of the image. In contrast, the outer regions store high-frequency data, contributing to edge definition and fine details (see Figure 2.1.6) [30] . Once k-space is fully sampled, the data can be transformed into the spatial domain to produce an image of the scanned tissue or organ via Fourier Transform [18] .

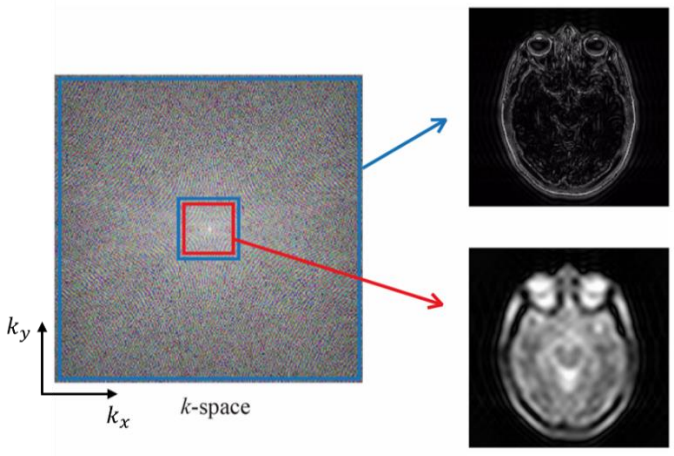


Figure 2.1.6 The k-space and corresponding image data from [31] .

ky is the phase encoding direction used to localize signals along the vertical direction. kx is the readout or frequency encoding direction summed to localize signals along the horizontal direction.

2.1.2 Structural MRI

In neuroimaging, structural MR images are typically used to show the gross anatomy of the brain: that is, mainly grey matter (GM) white matter (WM), and CSF. In a clinical setting these are the most commonly used images, and a radiologist typically inspects them visually, looking for tumors, pathological lesions, and anatomical deformations [19] .

The reason for having multiple types of structural images in common usage is that each type highlights different aspects of the tissues present, thus providing a valuable way of investigating the anatomy in vivo. T1-weighted and T2-weighted images are most common varieties of structural image [17] (see Figure 2.1.6). T1-weighted images highlight fat and subacute hemorrhage, offering excellent anatomical details. They are particularly useful for visualizing the brain structure, where they can distinguish between grey matter and white matter; T2-weighted images highlight fluid and edema, making them useful for detecting lesions, tumors, and areas of inflammation [29] .

Artifacts in structural MRI

MRI artifacts originate from numerous sources, such as MR hardware, room shielding, patient motion, tissue heterogeneity, foreign bodies, sampling resolution, and errors in k-space [32] . In structural imaging the most common artifacts (in Figure 2.1.7) are motion-induced artifacts, bias field or radio frequency inhomogeneity, ghosting, gradient nonlinearity distortion, and wrap around.

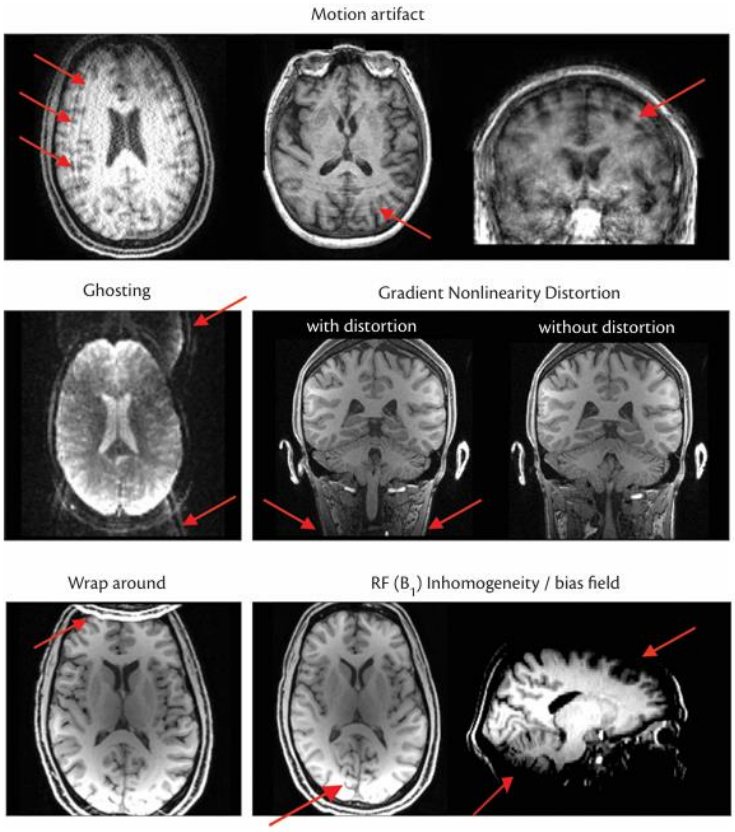


Figure 2.1.7 Typical artifacts from structural MRI images from [33] .

The top row shows motion-induced artifacts, characterized by the appearance of additional edges or ripples. The middle row illustrates ghosting, where a shifted and wrapped version of the image is superimposed on the original, alongside gradient nonlinearity distortion, which causes the neck to appear compressed horizontally towards the bottom of the image compared to an undistorted image. The bottom row depicts wrap-around artifacts, where the back of the head appears at the front, causing the signal of scalp to overlap with brain tissue, and bias field (or RF inhomogeneity), marked by a darkening at the top and bottom of the brain.

2.1.3 Functional MRI

Functional MRI assesses and maps brain activities by detecting changes in blood flow related to neural activity [38] . Specifically, it is sensitive to the hemoglobin, a blood component that behaves differently in its oxygenated and deoxygenated forms due to their distinct magnetic properties. This difference affects the local magnetic fields, primarily influenced by the levels of deoxygenated hemoglobin. These alterations create small, yet measurable changes in the MRI signal [33] .

When a brain region is more active, it consumes more oxygen, and the local cerebral blood flow increases, a relationship known as neurovascular coupling. fMRI

leverages this by measuring changes in the blood-oxygen-level-dependent (BOLD) effect (see Figure 2.1.8).

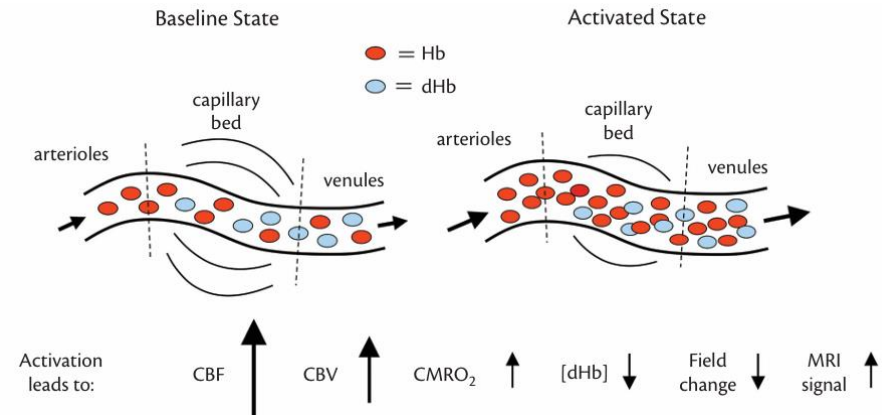


Figure 2.1.8 Illustration of the BOLD effect from [34] .

When local neuronal activity increases (transitioning from the baseline state on the left to the activated state on the right), cerebral blood flow (CBF) and cerebral blood volume (CBV) rise significantly. Oxygen consumption (CMRO2) also increases, but to a less extent, leading to a relative increase in oxygenated hemoglobin (Hb) compared to deoxygenated hemoglobin (dHb). This shift reduces the concentration of deoxyhemoglobin, which in turn decreases the magnetic field inhomogeneities it causes, resulting in an enhanced MRI signal.

There are two types of functional experiments, one involving stimuli and tasks, known generally as task fMRI, and the other conducted without explicit stimuli, known generally as resting-state fMRI. Task-based experiments are used to study the nature and location of specific processes in the brain, whereas resting experiments are primarily used to study functional connectivity between different regions of the brain.

Artifacts in functional MRI

As mentioned before, fast imaging is necessary for fMRI but brings with it several trade-offs, such as limited spatial resolution, low tissue contrast, geometric distortion, and signal loss [38] . These come in addition to the artifacts that can occur for structural MRI, especially motion, which may cause substantial signal changes. In fact, this is a major confound in fMRI, and specialized motion correction methods exist both in preprocessing and in the subsequent statistical analysis methods that aim to compensate for it as much as possible. Thus, methods for correcting the remaining geometric distortions and for compensating for the signal loss (both due to B_0 inhomogeneities) are required to be applied in the analysis of fMRI.

Due to the limited resolution and tissue contrast in these functional images, it is often difficult to identify in them detailed anatomical structures directly. It is even harder to align or register together with precision functional images that come from different subjects with different anatomies. For this reason, it is important to acquire a structural image for each subject, as this allows for greater precision in locating anatomical regions or borders as well as for greatly improved registration accuracy between subjects, which is important for having good statistical power in a group analysis.

2.2 Perfusion imaging using Arterial Spin Labelling MRI

2.2.1 Quantification of perfusion

Perfusion [39] [40] is the process by which blood delivers nutrients to tissues and removes wastes from tissues. Cerebral blood flow (CBF) is a quantitative measure of perfusion in the brain in unit of ml of blood per 100g of tissue per minute. In the brain, typical values of CBF are 60 ml/100g/min and 20 ml/100g/min for grey matter and white matter respectively [34] . There are two main reasons to measure perfusion. First is to measure the CBF in the resting state and associated parameters of blood flow. Abnormal values or changes of CBF could imply that the brain is experiencing some diseases. Second is to study and investigate the perfusion changes related to neuronal activities, which is also associated with the hemodynamics but from different views like cellular metabolism or electrical activity.

The gold standard of brain perfusion imaging is ^{15}O positron emission tomography (PET) [41] , which is performed by intravenous injection of $H_2{}^{15}O$ or inhalation of $C^{15}O_2$. However, the use of ionising radiation, the required access to on-site cyclotrons, and the expense of radiopharmaceuticals raise health and economic concerns for patients. As an alternative, the methods of CBF measurement using MRI generally includes the injection of an exogenous contrast agent, as in Dynamic Susceptibility Contrast-enhanced (DSC) MRI [42] and Dynamic Contrast-Enhanced (DCE) MRI [43] , or using an endogenous tracer, such as blood water in Arterial Spin Labelling (ASL) [54] . Specifically, ASL quantitatively measures cerebral perfusion, by taking advantage of using the magnetically labelled blood itself as an

endogenous tracer. Therefore, ASL has been extensively performed in the research area, and sporadically applied in clinic [44] .

2.2.2 Principles of Arterial Spin Labelling

Arterial Spin Labelling is unique in that it provides the only truly non-invasive method for measuring perfusion using blood water as the tracer [1] [34]. This uniqueness arises from the way the tracer is produced: radiofrequency fields are primarily used to invert the magnetic state of hydrogen nuclei into the opposite direction, typically at the neck as the blood flows into the brain. An ASL signal image typically involves acquiring a pair of images: a "label" image and a "control" image. Each image consists of a labelling stage, a waiting stage, and a readout stage. By introducing a delay (Post-Labelling Delay(PLD)) between the labelling and image acquisition, labelled blood is allowed to reach the capillaries, where it generates a perfusion signal and the image can be acquired which are called “labelled” image. In addition to labelled images, the separate “control” image is acquired without labelling arterial blood. The difference in signal between the control and labelled images quantifies the amount of labelled blood delivered to the tissue by perfusion. The difference between label and control images is typically around 1-2% compared to the signal, since the volume of labelled water delivered to the tissue is much smaller than the total amount of water in the tissue [1] . Therefore, it is common to use a low resolution, i.e., around 3 mm (to accumulate more signals from protons), and average multiple repeated volumes to achieve a better overall SNR which in turn trades the time of acquisition.

2.2.3 General principle of Arterial Spin Labelling

A basic ASL experiment generally consists of three steps: labelling, waiting, and imaging. This section tends to provide the mechanism of ASL with key concepts.

2.2.3.1 Labelling

As mentioned before, ASL labels blood-water as an endogenous tracer. To create the label, the magnetic properties of the hydrogen nuclei in the water are focused on. The hydrogen nuclei have their own magnetization including the magnitudes and the directions. When a strong magnetic field is placed, the directions of them can be aligned to the applied field. Therefore, hydrogen nuclei with the inversion of directions are created as a label. Three main labelling techniques commonly used for ASL are continuous Arterial Spin Labelling (cASL), pulsed Arterial Spin Labelling (pASL) and pseudo continuous Arterial Spin Labelling (pCASL) [34] [52] [54] . A simple process shown in Figure 2.2.1 includes these basic steps and the details are explained in following sections.

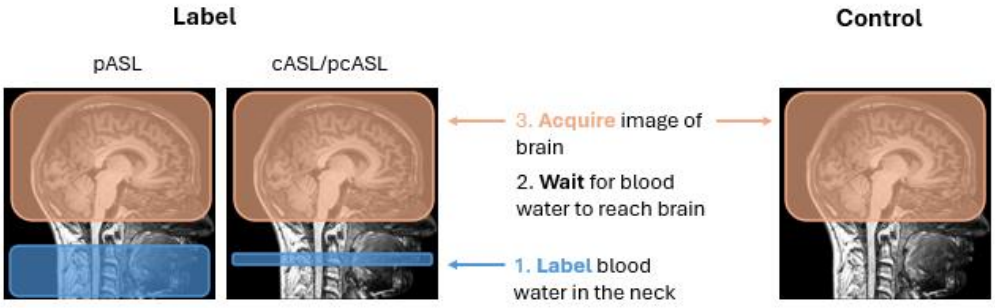


Figure 2.2.1 The process of ASL perfusion image acquisition.

The process requires a pair of images: one with labelling blood water (label) and the other without labelling (control). Only blood water in label images were labelled.

cASL

Continuous ASL (cASL) employs a continuous labelling scheme to magnetically invert arterial blood water spins. This is achieved by applying a continuous radiofrequency (RF) field in the presence of a magnetic field gradient (to induce flow-driven adiabatic inversions of blood spins [55]) across a proximal slice of the brain-feeding arteries, and the slice is known as the labelling plane.

As blood flows from below the neck toward the labelling plane, the magnetic field gradient causes the spins to become progressively less off-resonant with the RF pulse. When the spins reach the labelling plane and their off-resonance drops to zero due to the field gradient, the RF pulse would invert the magnetization of blood water.

This creates a continuous supply of labelled blood water flowing into the brain (see Figure 2.2.2).

cASL provides a consistent and predictable labelling pattern over time, and can achieve labelling efficiencies exceeding 90%, making it highly effective for perfusion imaging. However, the continuous application of RF pulses and magnetic field gradients requires specialized hardware, which is not typically available on commercial MRI scanners [54] . This limits the widespread adoption of cASL in clinical settings. Furthermore, the labelling efficiency of cASL is highly dependent on local blood velocity [53] . Because the labelling duration is finite, variations in blood flow speed (e.g., due to vascular disease or cardiac pulsation) can lead to inconsistent inversion of spins, reducing labelling efficiency and perfusion signal quality.

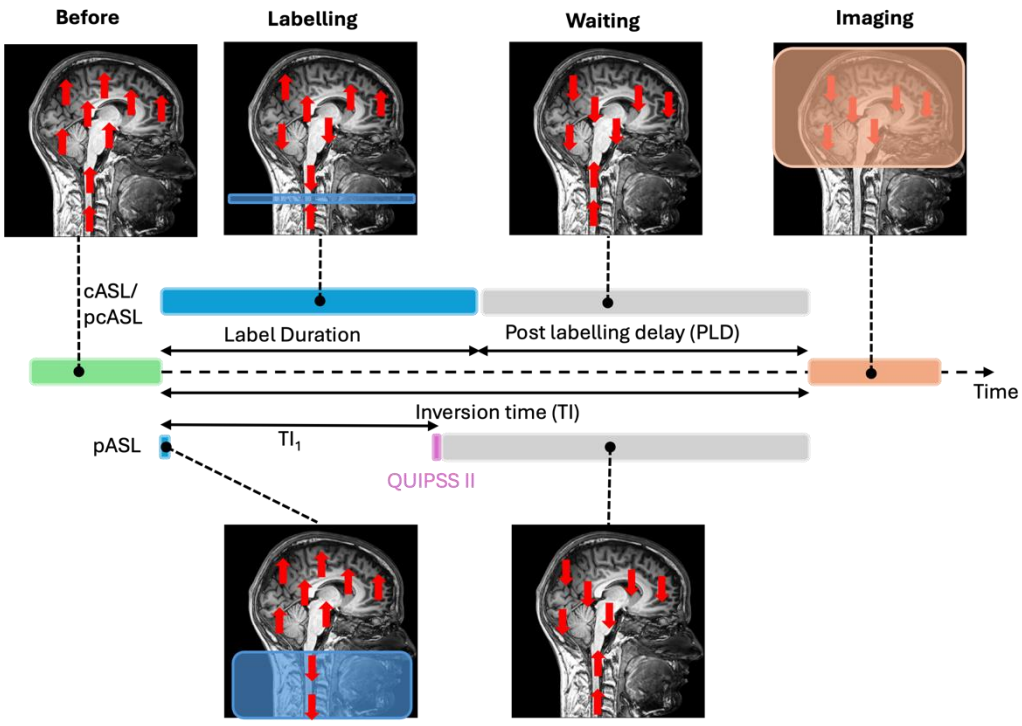


Figure 2.2.2 The typical processes for cASL, pASL, and pCASL acquisitions.

Different phases of ASL image acquisition were indicated with different colours in the middle timeline: before labelling (green), label duration (blue), saturation pulse (purple), post labelling delay (grey), and acquisition (orange).

pASL

Pulsed ASL (pASL) inverts the magnetization of blood water across a thicker region of space, which contains the brain-feeding arteries (see Figure 2.2.2). However, a

difference compared to cASL is that the labelling is nearly instantaneous (around 10 ms). In this case, labelling duration is unknown since it depends on the velocity of blood in the arteries within the specified region. Longer labelling durations require a larger spatial region for tagging, but this can reduce inversion efficiency for blood further from the center of the imaging slab due to the decreasing strength of the RF pulse with distance. In practice, it is hard to remain label durations for 1 s more. To achieve a constant time of labelling for pASL, methods such as QUIPSS I and QUIPSS II [56] with Q2TIPS [57] were commonly applied. QUIPSS I applies a saturation pulse to the imaging slice at time TI_1 after the inversion of the tagging region, creating a well-defined leading edge of the labelled slab. After a delay of ΔTI, blood starts to perfuse into the imaging slice, and the labelled image is acquired at $TI_2 = TI_1 + \Delta TI$. QUIPSS II, on the other hand, applies a saturation pulse to the labelling slice at the time TI_1, just as tagged blood exits the labelling region, producing a sharply defined bolus with time width TI_1 (see Figure 2.2.2). The Q2TIPS method improves on this by replacing the saturation pulse with a series of thin-slice periodic saturations at the distal end of the labelling region to reduce sensitivity to B_1 inhomogeneity and enhance the slice profile.

In contrast to cASL, pASL is much easier to implement with existing MRI hardware, as it does not require continuous RF pulses or field gradients, minimizing power deposition in tissue [34] . The short labelling duration also makes pASL relatively insensitive to magnetization transfer effects. However, the signal from more proximal regions of the imaging slab diminishes due to longer ATT and T1 relaxation, limiting the SNR of the pASL method.

pCASL

Pseudo-continuous ASL (pCASL) is also known as pulsed-continuous ASL which has a degree of similarity with cASL and can be interpreted in a similar way. It is currently the recommended method for clinical applications according to the ASL White Paper. pCASL resembles cASL in that it uses a constant labelling plane to invert the flowing spins as they pass through (Figure 2.2.2). However, in this case, the long label duration of cASL radiofrequency pulse and gradient are broken up into a series of repeating short radiofrequency pulses and associated gradients of pCASL,

which allows standard clinical scanners to create the labelling. The short RF pulses that can be controlled as identical ones and be synchronized to maintain balance in the labelling process. It is found that pCASL provides a 50% increase in SNR compared to pASL, and 18% higher in labelling efficiency compared to cASL [58] .

2.2.3.2 Waiting

After blood water is labelled, it requires some time to reach the brain tissue before the perfusion signal can be acquired. This waiting time varies slightly depending on the ASL method used (as shown in Figure 2.2.2). A fixed period of labelling duration of 1500–2000 ms is commonly applied for cASL and pCASL while blood has been labelled and started to flow toward the tissues [34] . In pASL, the inversion time (TI) is the period from the start of labelling to labelled blood reach the tissues. If a saturation pulse like QUIPSSII is applied, the time between the initial labelling and it is applied is known as TI_1, which is analogous to the labelling duration of cASL/pCASL. The PLD in pCASL is analogous to the quantity ($TI-TI_1$).

There are two main acquisition strategies based on how PLD is handled: single-PLD and multi-PLD. In a single-PLD protocol, a single fixed waiting time is used for all measurements, after all labelled blood reach tissues, and Arterial Transit Time (ATT) is the physiological parameter to represent the time taken for the labelled spins to reach the brain region of interest. This means the same delay between labelling and imaging is applied across all acquisitions, providing perfusion images that reflect the blood flow at that specific time point. However, to accommodate the longest ATT expected in a given group of subjects, a longer PLD must be chosen to ensure complete blood arrival in the tissue. This longer PLD reduces the ASL signal strength. While single-PLD is simpler to implement and less computationally intensive, it provides limited insight into the dynamic changes in blood flow over time. Additionally, it assumes that all blood arrives at the tissue simultaneously, which can lead to either under- or over-estimation of CBF, especially in regions with delayed or prolonged ATT [34] . In contrast, multi-PLD involves acquiring multiple images with varying post-labelling delays. By using different waiting times, this approach captures the dynamic evolution of blood flow at different time points, offering a more comprehensive view of perfusion. Multi-PLD enhances the accuracy

of CBF quantification by enabling more complex kinetic modelling of blood flow and tissue perfusion. It is less sensitive to differences in ATT across tissues and subjects compared to the single-PLD protocol. However, this technique requires longer scan times and more complex data analysis [24] .

2.2.3.3 Readout

After it is ensured that all labelled blood-water reach the tissue in the brain, fast imaging techniques are used to image the brain acquisition (known as readout). Various fast imaging techniques are available, such as Echo-planar imaging (EPI) [61] and Spiral [62] .

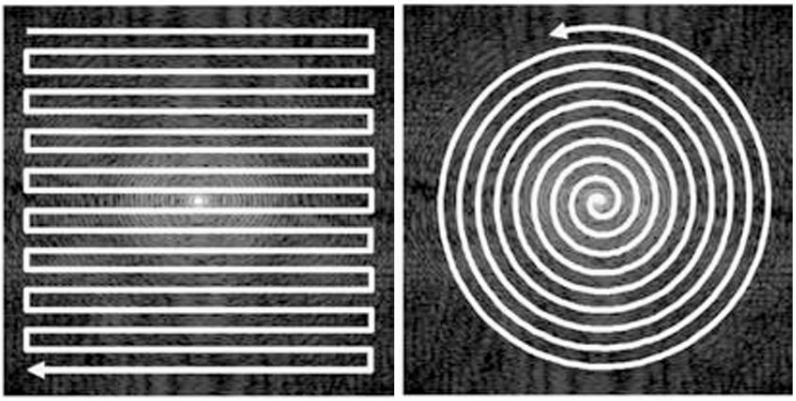


Figure 2.2.3 Examples of k-space trajectories. Left: rectilinear; Right:Spiral from [65] .

The EPI sequence scans k-space in a rectilinear (zigzag) pattern (see Figure 2.2.3), and its set of excitation pulses allows for rapid acquisition of each brain slice, typically within 50 ms. As a result, multiple lines or even the entire k-space can be sampled within a single TR, which is referred to as multi-shot EPI or single-shot EPI, respectively. A sequence of oscillating readout gradients follows, each accompanied by a brief, low-amplitude "blipped" phase encoding gradient (Figure 2.2.4). However, EPI suffers from signal dropout and distortion artifacts in regions with poor magnetic field homogeneity.

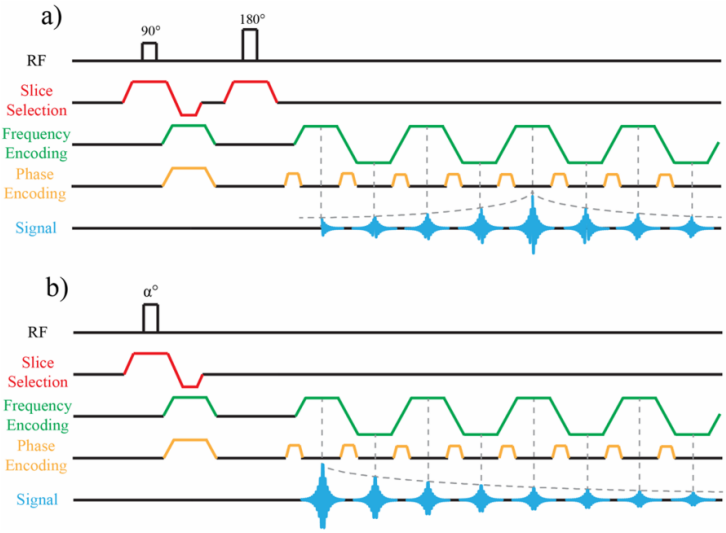


Figure 2.2.4 The Echo Planar Imaging (EPI) sequence diagram from [31] :(a) Spin-echo EPI – A 90° flip angle is applied, followed by a 180° refocusing pulse. (b) Gradient-echo EPI – The flip angle is variable (less than 90°).

3D Gradient and Spin Echo (GRASE) [63] combines elements of both gradient echo and spin echo techniques. This hybrid approach provides minimal signal dropout with reduced susceptibility artifacts compared to pure gradient echo methods. It offers a compromise between image quality and acquisition speed, making it suitable for a broader range of clinical applications.

Spiral imaging, on the other hand, employs rotating gradients to trace a spiral trajectory across k-space (Figure 2.2.3), moving from the center outward. Spiral readout techniques typically achieve a shorter minimum TE than EPI, which reduces susceptibility to motion artifacts. However, spiral imaging is traversed continuously without refocusing gradients, and thus off-resonance could accumulate due to prolonged readout, potentially leading to image blurring. 3D rapid acquisition relaxation enhanced (RARE) Stack of Spirals offers better efficiency by oversampling at the center of k-space, thus provides shorter TE and minimal signal dropout, and exhibits less sensitivity to motion artifacts compared to GRASE [66] , but it could introduce in-plane blurring due to longer readout times.

A higher SNR can be achieved using 3D MRI with a thick volume selectively excited by the RF pulse [1] , as the 3D imaging methods using a thicker bandwidth, instead of exciting thin slices, could mitigate the off-resonance effects by averaging more

local spins over more contiguous slices and shorten the time for off-resonance shifts accumulation [14] . Thus, 3D readout techniques are recommended for ASL [1] as they acquire the entire image volume in a single TR, providing an optimal SNR while incorporating background suppression (BS) [57] [60] . These methods enhance ASL signal measurement and are relatively insensitive to field inhomogeneity. Background suppression further improves SNR by minimizing static tissue signal during readout, increasing the contrast between labelled blood and surrounding tissue. This, in turn, enhances perfusion detectability and overall image quality.

2.3 ASL MRI Processing

2.3.1 Motion correction

During the process of collecting a series of 3D volumes, it is necessary to compensate for any subject motion which occurs between acquisitions. This kind of motion is likely to occur in acquisitions of functional and diffusion images which requires fast or continuous acquisitions of hundreds 3D volumes, but it may also take place in the acquisition of multiple repeats for a structural sequence [63] .

ASL is also susceptible to motion, as the label and control images are acquired multiple times. Subtle head motion can lead to significant voxel difference when calculating the subtraction between label-control images, which is especially visible in voxels around the edge of brain due to the substantial contrast between brain and air [34] . Thus, it is required to apply motion correction before the subtraction of label-control images. The main method to correct head motion is to apply the rigid-body registration between individual volumes with a consistent reference image, which can be the first, middle, or an average image in the series. However, a drawback of motion correction is that subtraction artifacts can be induced due to the interpolation from registration-based motion correction [64] .

2.3.2 Distortion correction

ASL data acquired using EPI-based techniques will commonly suffer from distortions [34] ,[67] . EPI acquisitions are particularly sensitive to imperfections in

the B_0 field, leading to significant geometric distortions and localized signal loss (or dropout). Additionally, EPI images tend to have lower spatial resolution and may suffer from poor tissue contrast. As a result, accurately aligning them with structural images, even when the same anatomy is present, can be challenging. The most problematic distortion is the geometric one caused by B_0 field inhomogeneities, which are primarily due to air-filled sinuses and cavities in the skull. This issue is especially problematic because it results in localized changes in the brain, often accompanied by signal loss. To correct these distortions, two main approaches can be used to obtain information about B_0 inhomogeneities [34] .

One involves acquiring two undistorted gradient-echo images (like a structural MRI sequence, so it is undistorted) with different echo times (so signal is sensitive to B_0 inhomogeneities) and phase information can be calculated for the B_0 field at each voxel, as the difference in phase can give a rate of change of phase that divided by the difference in echo times is equal to the value of B_0 field; this is known as the gradient-echo field map. Once the geometric distortion is calculated, it could be incorporated, as a spatial transformation, into a rigid-body registration (with EPI images) for distortion correction.

The other one involves acquiring two distorted images using spin-echo EPI with opposite phase-encode directions, a technique often referred to as blip-up-blip-down acquisition or phase-encode reversed acquisition. The distortion in the two images can be calculated because the direction of the distortion is reversed in the two images. The "halfway" image, representing the undistorted state, can then be derived by registering the two distorted images. Thus, the calculated spatial transformation can be used to correct distortion as a separate step in the preprocessing pipeline before registration.

2.3.3 Registration

In MRI, the primary goal of registration is to align anatomical structures between different images, especially for group studies. Linear transformations [68] are commonly used when the brain anatomy remains unchanged. For example, if the head moves during an ASL scan, only translations (shifts) and rotations are required.

These transformations are captured by rigid-body registration, which involves 6 degrees of freedom (DOF)—three for rotations and three for translations.

Another common linear transformation is the affine transformation, which has 12 DOF (six from rigid-body transformations, plus three for scaling and three for shearing). While affine transformations can change the shape and the size of an image, they do not represent realistic brain motion. However, they are helpful for eddy-current distortion correction and initializing nonlinear transformations.

When there is a significant anatomical difference or local geometric changes, linear transformations are insufficient. In such cases, nonlinear transformations are necessary to align images accurately. Nonlinear transformations [69] , often referred to as warps, involve more than 12 DOF and can range from thousands to millions of DOF. These transformations describe local changes in geometry, such as displacements at specific grid points, which account for how that point and its surrounding regions are shifted [71] .

Like many other fMRI data, it is often required to convert low resolution ASL data from a subject to different resolutions and orientation. A typical example is the need to convert a group of MRI images into a common space, such as MNI152 standard space [70] , for group analysis. The conventional processes of reasonable registrations for ASL data involve: rigid registration from ASL native space to the structural space, and non-rigid registration from the structural space to the standard space (see Figure 2.3.1).

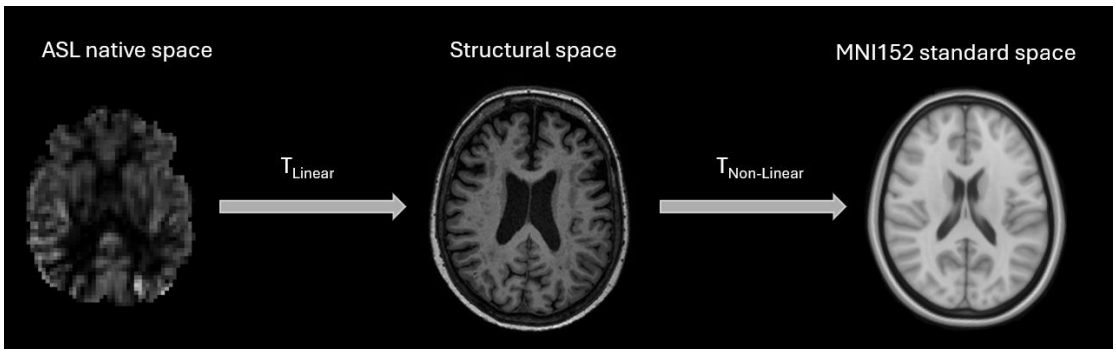


Figure 2.3.1 The process of transforming a CBF map to structural space and then to standard space.

T_s is the transformation of linear rigid-body registration which align the CBF map to the T1-weighted structural image in the same subject. T_t is nonlinear registration to estimate the transformation from T1-weighted structural image to the standard space (the MNI152 space in this case). Interpolation is used in these transformations to obtain the perfusion images in the same resolution with the structural image or template image.

2.3.4 Quantification

To obtain quantitative measurements of perfusion using ASL, a comprehensive analysis process is required, which typically involves three key components: subtraction, kinetic modelling, and calibration.

The label and control images include contributions from static tissue water, but only the label image contains the inverted signal from labelled blood water, resulting in a negative contribution to the perfusion signal. The subtraction is to isolate the signal of labelled blood water in the tissue, which is achieved through the pairwise subtraction between the label image and the control image. Therefore, the static tissue signal is effectively removed, leaving behind a perfusion-weighted image that exclusively reflects the signal from labelled blood water. Assuming the blood water magnetisation is completely relaxed within the control image (M_{0a}), the ASL signal difference (ΔM) between the longitudinal magnetization of the label and control images by local perfusion (f) during time delay (TI) can be expressed in Equation 2.3.1.

$$\Delta M = M_{\text{control}} - M_{\text{label}} \;=\; 2M_{0a} \otimes f \otimes TI \qquad \text{Equation 2.3.1}$$

Where $\otimes$ indicates the convolution operation, since the perfusion-weighted signal reflects the amount of labelled blood water that has accumulated in the tissue during PLD. To translate the perfusion-weighted signal into quantitative perfusion values, a tracer kinetic model is needed be applied. A widely used model for ASL is the general kinetic model proposed by Buxton et al. [72] , which has become a standard for single-PLD ASL data. This model is based on three key assumptions:

- The arrival of labelled blood at a specific voxel is assumed to follow a uniform plug flow. This means that before the initial transit delay Δt, no labelled blood reaches the voxel. Between time $t = \Delta t$ and $t = \tau + \Delta t$ the blood arriving is uniformly labelled, with that τ is the label duration. After $t \geq \tau + \Delta t$, the arriving blood is again unlabelled. For pCASL, this requires PLD > ATT, and for QUIPSS II pASL, it needs ($TI - TI_1$)>ATT. This ensures sufficient time for bolus arrival. Notably, this simplification neglects dispersion, a phenomenon arising from physiological factors such as transit

through progressively narrower vessels, flow through the branching vascular network, and cardiac pulsation. The effect could collectively smooth the AIF, deviating from the idealized sharp arrival assumed in the plug-flow model.

- The kinetics of water exchange between blood and tissue are assumed to be modelled using a single-compartment model. This approach assumes that even if sub-compartments exist within the tissue, water molecules exchange between them so rapidly that their concentration ratios remain constant over time and equal to the brain/blood water partition coefficient λ, despite changes in the total tissue concentration, which leads to the "well-mixed" assumption, i.e. the label concentration in capillary blood and tissue equilibrates sufficiently rapidly in the voxel. Consequently, the concentration at the venous end of the capillary bed will match the label concentration distributed throughout the voxel. Under the assumption, the dominant mechanism for signal loss is T1-relaxation (the decay of magnetization due to longitudinal relaxation), rather than physiological processes such as vascular outflow.

- After the inversion pulse, relaxation is assumed to be governed by blood T1 relaxation (T_{1b}), and after labelled blood, the magnetization decreases with tissue T1 relaxation (T_{1t}). This assumption implies that water is instantly extracted from the vascular space upon reaching the voxel. However, in fact the labelled water may remain in the blood vessels for some time after arrival, as the tagged blood continues to flow through the vascular network toward the capillary bed.

These assumptions simplify the kinetic model, making it computationally tractable while still providing accurate estimates of CBF in most physiological scenarios. This model explicitly accounts for the three physiological processes governing the behaviour of labelled blood water in tissue: arterial input function (AIF), residue function, and magnetization relaxation function. The AIF $a(t)$ quantifies the concentration of labelled spins arriving in the voxel via arterial blood at time t. However, once delivered, a fraction of these spins is removed from the voxel through venous clearance or further perfusion. The residue function $r(t)$ describes the proportion of labelled spins remaining in the voxel at time t after their initial

delivery. Concurrently, the inverted magnetization of labelled blood water undergoes longitudinal relaxation at a rate determined by the T1 of arterial blood. The magnetization relaxation function $m(t)$ accounts for the decay of residual magnetization over time t, reflecting the progressive loss of signal due to T1 relaxation. Under the assumption that off-resonance effects and magnetization transfer [245] are identical in both label and control images, the time-dependent ASL signal difference $\Delta M(t)$ becomes directly proportional to the change of perfusion magnetization. According to $\Delta M = M_{\text{control}} - M_{\text{label}} = 2M_{0a} \otimes f \otimes TI$ Equation 2.3.1, this can be expressed as:

$$\Delta M(t) = 2M_{0a} f\{a(t) \otimes [r(t)m(t)]\} \qquad \text{Equation 2.3.2}$$

where M_{0a} represents the equilibrium magnetisation of arterial blood, f is the perfusion in ml/g/s.

To be detailed, the AIF describes the time course of delivery of the labelled blood water to the part of the brain being imaged, is mathematically described as:

$$a(t) = \begin{cases} 0 & t < \Delta t \\ 2M_{0a} e^{-t/T_{1b}} \quad (\text{pASL}) & \Delta t \le t < \tau + \Delta t \\ 2M_{0a} e^{-\Delta t/T_{1b}} \quad (\text{pCASL}) & \\ 0 & t \ge \tau + \Delta t \end{cases} \qquad \text{Equation 2.3.3}$$

where M_{0a} is the magnetization of the labelled arterial blood at the neck, T_{1b} is the T1 of the arterial blood, τ is the label duration, and Δt refers to the ATT.

Once the labelled blood water reaches the tissue through the vasculature, it remains there for a period determined by either the time it takes for the labelled water to leave the voxel or the rate at which the label decays due to T1 relaxation. The residue function, as shown in Equation 2.3.4, quantifies the proportion of labelled blood-water still present in each voxel at a given time point.

$$R(t) = r(t)m(t) \qquad \text{Equation 2.3.4}$$

with $r(t) = e^{-tf/\lambda}$ represents the outflow effect, and by assuming a constant concentration ratio between tissue and venous areas, λ of 0.9 is the partition coefficient of tissue/blood water. $m(t) = e^{-t/T_{1t}}$ capturing the magnetization decay, where T_{1t} is the T1 of the tissue in the voxel.

The kinetic model for the signal difference in magnetization between label and control conditions can be formulated:

for pCASL/cASL

$$\Delta M(t) = \begin{cases} 0 & t < \Delta t \\ 2M_{0a} f T_{1app} e^{-\Delta t/T_{1b}} (1 - e^{-(t-\Delta t)/T_{1app}}) & \Delta t \le t < \Delta t + \tau \\ 2M_{0a} f T_{1app} e^{-\Delta t/T_{1b}} e^{-(t-\Delta t-\tau)/T_{1app}} \left(1 - e^{-\tau/T_{1app}}\right) & \Delta t + \tau \le t \end{cases}$$

Equation 2.3.5

and for pASL

$$\Delta M(t) = \begin{cases} 0 & t < \Delta t \\ 2M_{0a} f \mathrm{e}^{-\mathrm{t}/\mathrm{T}_{1\mathrm{b}}} \frac{\mathrm{e}^{\mathrm{Rt}}(\mathrm{e}^{-\mathrm{R}\Delta\mathrm{t}} - \mathrm{e}^{-\mathrm{Rt}})}{R} & \Delta t \le t < \Delta t + \tau \\ 2M_{0a} f \mathrm{e}^{-\mathrm{t}/\mathrm{T}_{1\mathrm{b}}} \frac{\mathrm{e}^{\mathrm{Rt}}(\mathrm{e}^{-\mathrm{R}\Delta\mathrm{t}} - \mathrm{e}^{-\mathrm{R}(\Delta\mathrm{t}+\tau)})}{R} & \Delta\mathrm{t} + \tau \le \mathrm{t} \end{cases}$$

Equation 2.3.6

Where T_{1app} is the apparent tissue relaxation time, associated with the decrease of longitudinal magnetisation additionally caused by labelled blood flow, whose magnitude depends on the relaxation time of tissue (T_{1t}) and blood flow : $\frac{1}{T_{1app}} = \frac{1}{T_{1t}} + \frac{f}{\lambda}$, and $R = \frac{1}{T_{1b}} - \frac{1}{T_{1app}}$.

Thus, a kinetic model curve of ASL signal difference $\Delta M(t)$ can be expressed in Figure 2.3.2.

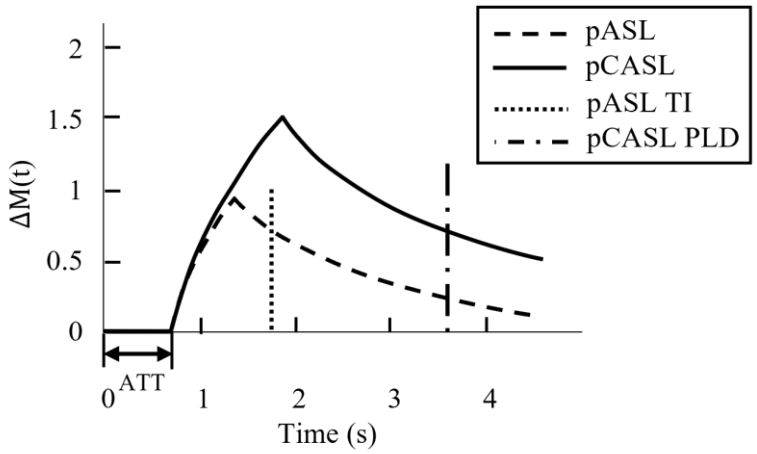


Figure 2.3.2 An example of the general kinetic model curve in a voxel for pASL and pCASL.

The parameters for pcASL and pASL uses the recommendations in the consensus paper: a bolus duration of 1.8 s for pcASL and 0.8 s for pASL;PLD of 1.8s for pCASL and TI of 1.8s for pASL. ATT=Arterial Transit Time, the time required for labelled blood to traverse arteries and reach capillaries.

To quantify the tracer concentration in the unit of CBF, the signal intensities, which typically vary across individuals, as well as MRI-related factors such as the main magnetic field, need to be scaled [34] . This process is known as calibration. There are two main opinions on how to determine the value: one is the voxel-wise approach and the other one is reference region approach.

In voxel-wise approach, the values for each voxel are estimated based on a separately acquired PD-weighted image. The rationale behind this method is that it can automatically correct for voxel-wise variations in intensity, such as coil sensitivity, and typically involves simpler calculations. The PD-weighted image is typically obtained from a separate calibration scan acquired with a long TR to ensure proton density weighting. The other approach involves taking the mean value over a reference region as a single global value. The argument for the reference region approach is that different values in each voxel is not acceptable since it is expected to have a single value of magnetization for the arterial blood as it is passing through the labelling plane.

In the consensus paper, the voxel-wise approach is recommended in the consensus paper [1] for general purpose calibration. Furthermore, the general kinetic model was further simplified from the general kinetic model by assuming that the T1 of tissue (T_{1t}) is equivalent to the T1 of blood (T_{1b}), and no outflow condition due to large water pool, and. As a result, the absolute quantification of CBF can be derived using a set of label-control paired ASL images along with a PD weighted image, as described by the following Equation 2.3.7 and Equation 2.3.8:

using pCASL/cASL:

$$CBF = \frac{6000\lambda(S_{control}-S_{label})e^{\frac{PLD}{T_{1b}}}}{2\alpha T_{1b}S_{PD}(1-e^{-\frac{\tau}{T_{1b}}})}\,(\mathrm{ml/100g/min}) \qquad \text{Equation 2.3.7}$$

and using QUIPSS II pASL:

$$CBF = \frac{6000\lambda(S_{control}-S_{label})e^{\frac{TI}{T_{1b}}}}{2\alpha TI_1 S_{PD}}(\text{ml/100g/min}) \qquad \text{Equation 2.3.8}$$

Where $S_{control}$ and S_{label} are the time-averaged signal intensities of the control and label images, respectively. SI_{PD} refers to the signal intensity of a proton density-weighted image. α is the labelling efficiency of 0.98 for pASL and 0.85 for pCASL. $T_{1\,b}$ decay is time constant of 1.65 s. TI is the inversion time, and TI_1 is the time intervals for the QUIPSS II tagging pulse of 800 ms. PLD is the post-labelling delay of 1800 ms. The factor of 6000 converts the units from ml/g/s of perfusion to the customary physiological units of ml/100g/min.

The general kinetic model uses a single-PLD, chosen to allow labelled blood sufficient time to reach the tissue. While simple and fast, it risks underestimating CBF in regions with delayed ATT (Figure 2.3.3).

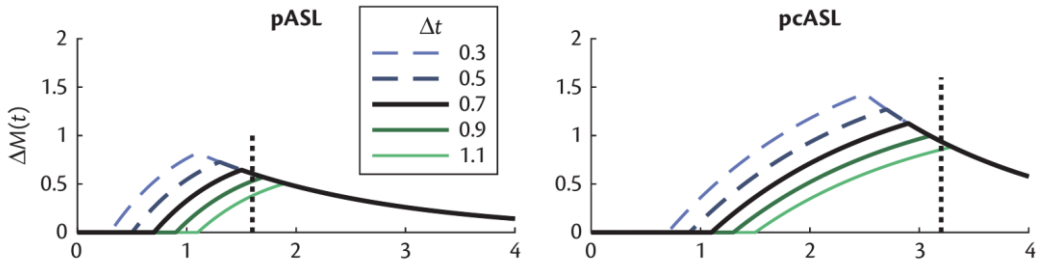


Figure 2.3.3 Variation in the ASL signal at a voxel in the brain, in response to changes in ATT, as well as for both pcASL and pASL from [34] .When ATT is 0.9 or 1.1 (>PLD), labelled water is not completely reach tissue.

To mitigate this, consensus PLDs are set conservatively long, or flow suppression techniques [60] are applied to suppress signals from fast-moving arterial blood. On the other hand, the multi-PLD method samples the kinetic curve at multiple delays (Figure 2.3.4), enabling simultaneous estimation of CBF, ATT, and aBV (arterial blood volume).

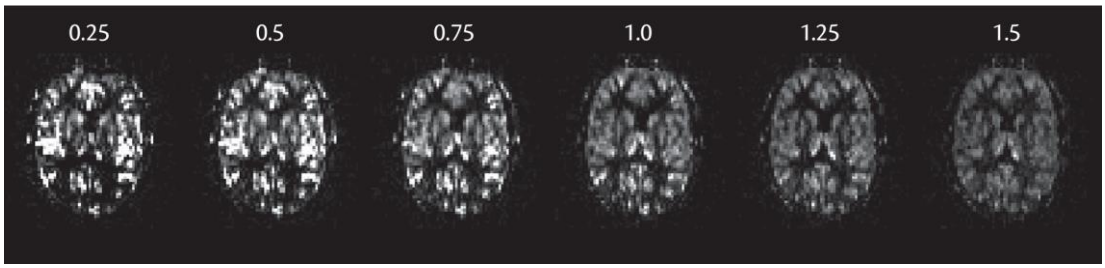


Figure 2.3.4 An example of the label-control difference images at various PLD values (in seconds), after averaging the repeated measurements for each PLD from [34] .

By varying the PLD, multi-PLD ASL acquires a series of images that capture different stages of the arrival and washout of labelled blood water in the tissue. Tracking the arrival of labelled blood at different tissues over time enables the estimation of ATT. Furthermore, multi-PLD ASL data allow for distinguishing the contribution of aBV from tissue perfusion. aBV represents the amount of labelled blood within the arterial compartment of a voxel, and its influence is most prominent in areas with high vascular density, such as inferior brain regions near major arteries. Additionally, the aBV map derived from multi-PLD data highlights vascular structures, like the middle cerebral arteries, and can aid in differentiating arterial contamination from true perfusion signals [24] .

2.4 Partial volume effects

Partial volume effects (PVE) are a common issue in MRI imaging, leading to voxels containing a mixture of different tissue types, which can obscure the true representation of any single component and the sampling of continuous anatomical structures. This would affect the precise measurement of a specific type of tissue of interest. To ensure accurate quantification of MRI signals, it is essential to understand the implications of PVE and apply appropriate correction methods for this source of error.

2.4.1 Partial volume effects in ASL MRI

Partial volume (PV) effects can significantly impact the quantification of perfusion in ASL MRI [73]]. The ASL signal derived from the brain typically contains three

primary sources: grey matter, white matter, and cerebrospinal fluid. Ideally, the ASL signal from CSF should be negligible, as labelled blood water is not expected to reach the CSF [74] [75] . Within the context of perfusion MRI, GM is the main interest due to its critical functions in the central nervous system and the observed subtle variations in GM CBF associated with various diseases, for example, dementia [76] . However, the spatial resolution of ASL data is relatively low, typically between 3–5 mm, which often does not match the 2–4 mm thickness of GM in cortical areas [77] , thereby exacerbating partial volume effects (PVE). While all MRI modalities used in brain imaging are susceptible to PVE, these effects are particularly pronounced in perfusion imaging due to the significantly different perfusion characteristics of GM compared to WM, where a perfusion ratio of 3:1 is frequently assumed [34] . While this may not be problematic when visually assessing areas of hyper- or hypo-perfusion (e.g., in brain tumors or stroke), PVE can distort quantitative CBF measurements, especially when comparing individuals or detecting subtle changes. This is particularly pertinent in research involving patients with tissue atrophy [73] or neurodegenerative diseases, such as Parkinson's disease, where alterations in voxel tissue content might be mistaken for perfusion changes or masked by CBF differences between adjacent anatomical voxels.

2.4.2 Partial volume effects correction

It is becoming more widely acknowledged that addressing PVE is crucial for accurately quantifying grey matter CBF without interference from the partial volume of WM or CSF. Several partial volume effects correction (PVEc) methods have been developed to address partial volume effects in perfusion estimation from ASL data. These methods include the linear regression (LR) technique [7] and a spatially regularized approach that incorporates spatial priors within a bayesian inference framework for perfusion quantification [8] .

A general formulation of the ASL signal, accounting for contributions from both GM and WM, can be expressed as:

$$\Delta M(t) = P_{GM} \cdot \Delta M_{GM}(t) + P_{WM} \cdot \Delta M_{WM}(t) \qquad \text{Equation 2.4.1}$$

In this formulation, ΔM represents the difference in longitudinal magnetization between the ASL label and control images, P denotes the partial volume estimates for each tissue type, and t indicates the time elapsed since the start of the RF

inversion. It is assumed that the contribution from cerebrospinal fluid (CSF) is negligible, as no significant ASL difference magnetization is expected in the CSF. This approach is applicable to both single and multi-PLD ASL data.

Estimating the magnetization of either grey or white matter independently within a voxel is challenging due to the availability of only a single measurement of magnetization (ΔM). PVEc methods address this issue by incorporating information from neighboring voxels, with different strategies used to integrate this data into the estimation process.

The linear regression approach for PVE correction in ASL assumes that CBF in both GM and WM can be considered constant within a 2D regression kernel of size n × n, centered on the voxel where GM and WM values are required. Under this assumption, where GM and WM CBF values are constant across voxels within the kernel, the ASL difference signal in the voxels can be expressed in matrix form based on Equation (2.5.2).

$$\Delta M = P\Delta\overline{M} \qquad \text{Equation 2.4.2}$$

Where ΔM is a vector of length n^2 containing all the ASL difference values, P is n^2 x 2 matrix of GM and WM PV estimates and $\Delta\overline{M}$ is a vector with two entries representing the grey and white matter perfusion within the kernel. The numerical solution can be found using the following Equation 2.4.3, which minimizes the squared error and is mathematically equivalent to linear regression:

$$\Delta\overline{M} = (P^T \cdot P)^{-1} \cdot P^T \cdot \Delta M \qquad \text{Equation 2.4.3}$$

where $(P^T \cdot P)^{-1}$ is the pseudo-inverse matrix of P. This process is repeated for each voxel in the brain to generate a map of PV-corrected GM and WM ASL difference signals. The method can also be applied to the final quantified perfusion images [175]], and in principle, it can accommodate various kernel shapes, as long as they are defined in terms of whole voxels.

An alternative spatially regularized approach proposed by [8] uses the formulation in Equation 2.4.1 but incorporates spatial priors for both GM and WM within a variational bayesian inference framework. This method is based on the same principle as the linear regression approach but improves upon it by allowing spatial regularization to be adaptively guided by the data, rather than relying on a fixed

kernel size. The algorithm estimates both GM and WM contributions in each voxel, with each contribution influenced by a prior distribution derived from the neighboring voxels, referred to as the spatial prior [80] ,[81] . This process is iterative, with voxel values being updated in tandem with the spatial prior and repeated in cycles. By applying Bayes' theorem, the voxel estimates are informed by both the measured data and the prior, with the balance automatically adjusted, allowing the spatial regularization to adapt to the data. While initially demonstrated with multi-PLD ASL data, this approach is also applicable to single-PLD ASL.

2.5 Artifacts in ASL MRI

Artifacts are distortions or anomalies in an image that do not accurately represent the tissue or organ being scanned. In ASL MRI, artifacts can occur at different stages of sequence acquisition [82] , including during the tagging phase (magnetic labelling), the transit phase (the time between labelling and image acquisition), or the readout phase (image acquisition itself). Additionally, various physiological changes can also introduce artifacts, which radiologists must be mindful of when interpreting ASL images. This section provides an overview of common artifacts encountered in ASL MRI.

2.5.1 Artifacts arising during labelling

Ineffective labelling

To ensure that the distribution of ASL signal within brain tissue accurately interprets the pattern of cerebral blood flow, protons in arterial blood close to the imaging plane must be labelled efficiently and uniformly. There are two factors can influence the process: vessel tortuosity and susceptibility variation within the labelling region.

Inefficient labelling, resulting from these factors, can lead to ASL signal loss in the impacted vascular area. If the orientation of a target artery is not perpendicular to the labelling plane, it can reduce the efficiency of the inversion process by slowing down the blood flow. This reduced inversion efficiency may result in a weaker labelling of the arterial blood, ultimately diminishing the ASL signal. Additionally, severe vessel

tortuosity, where the artery crosses the labelling plane multiple times, can lead to multiple inversions of the intraluminal blood, thereby diminishing labelling efficiency, as shown in Figure 2.5.1. In fact, although tortuosity of the cervical internal carotid artery occurs frequently, it seldom results in labelling failure.

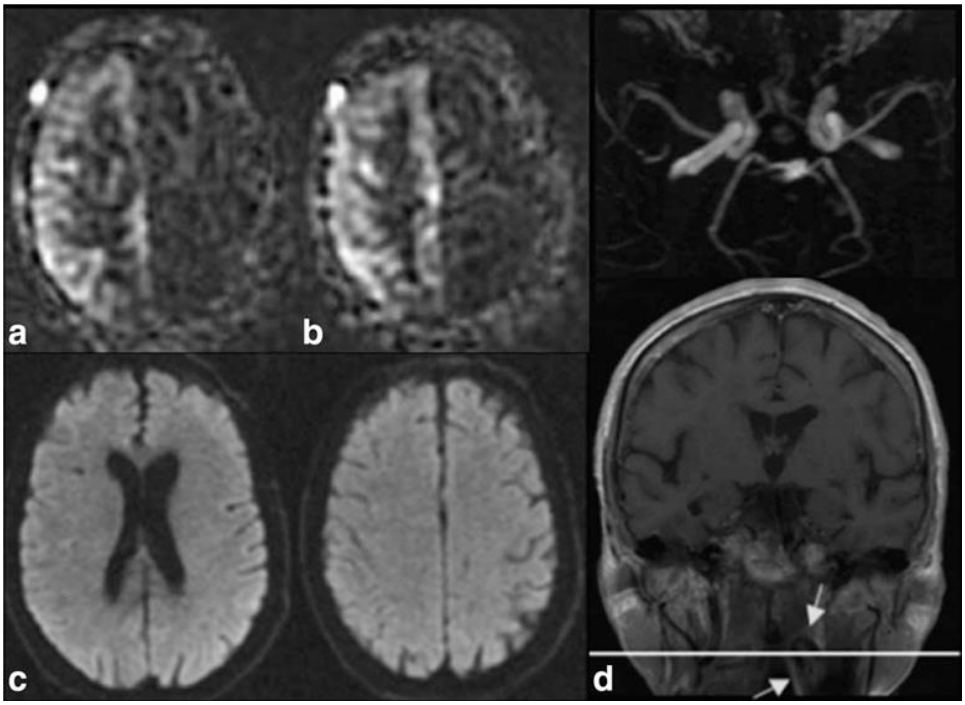


Figure 2.5.1 Poor labelling caused by vessel tortuosity from [82] .

On the ASL CBF maps (a), there is a noticeable absence of signal in the left internal carotid artery (ICA) territory. However, collapsed 3D time-of-flight MR angiography (TOF MRA) (b) reveals normal flow signal in the left ICA, and the DWI image (c) shows no acute infarction in the brain. (d) The distal ICA is visible on coronal post-contrast T1-weighted images. The right ICA (arrowed) shows significant tortuosity as it passes through the labelling plane (white line) twice, leading to multiple inversions and reduced labelling efficiency. This results in decreased signal in the right parietal lobe on the ASL CBF maps.

Susceptibility variations within the labelling plane can cause dephasing of arterial blood protons, disrupting the conditions needed for pseudo-continuous inversion and resulting in poor or absent labelling, as illustrated in Figure 2.5.2. Various factors can contribute to susceptibility variations, including metallic surgical hardware, dental fixtures, calcification, and air–tissue interfaces caused by the pneumatization of skull base and facial bones. Although these factors are common, the impact on labelling efficiency due to susceptibility effects is relatively rare. This is likely because the susceptibility gradients generated by these materials are generally too weak or distant to significantly change the magnetic flux density in the internal carotid and vertebral arteries.

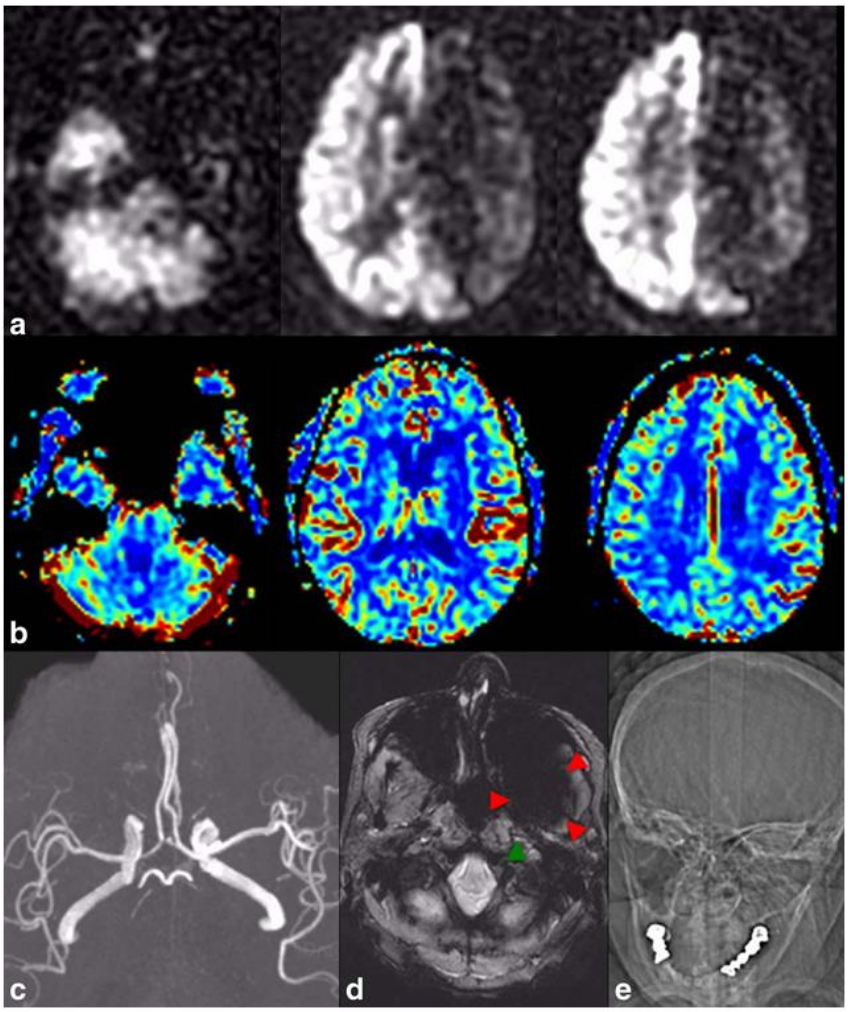


Figure 2.5.2 Poor labelling due to susceptibility variations in the labelling plane from [82] .

The ASL CBF maps (a) show a decreased signal in the left ICA territory, suggesting potential perfusion issues. However, the relative CBF maps (b) generated from DSC perfusion imaging in the same session shows the interhemispheric symmetry of blood flow, which is inconsistent with the ASL findings. Additionally, there is no steno-occlusive disease observed in the left ICA from TOF MRA(c). An axial T2-weighted gradient-echo image (d) at the level of the maxillary antra reveals an extensive area of signal loss (red arrowheads), indicating intravoxel dephasing due to susceptibility variations. A prior CT scout image (e) identifies the causative factor as dental hardware, specifically on the left side.

Cerebrospinal Fluid Labelling

Artifacts can also occur during the labelling period due to the inversion of water molecules located outside the intraarterial compartment. If these labelled nonarterial water molecules subsequently move into the imaging plane, they can contribute to artifacts. For example, water molecules in the cerebrospinal fluid can be tagged during the labelling phase and then migrate into the imaging plane, driven by pulsatile flow—particularly under conditions of hyperdynamic CSF flow (as shown in Figure 2.5.3). This migration often results in a distinctive high signal around the medulla, creating what is commonly referred to as a "ring-of-fire" appearance.

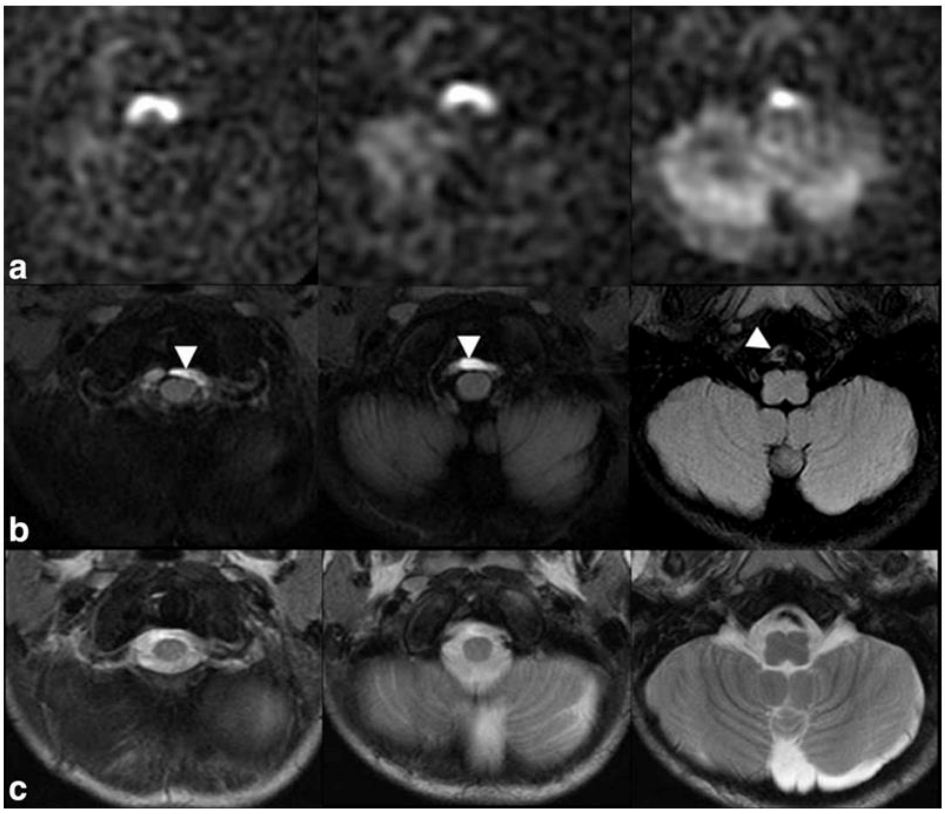


Figure 2.5.3 CSF labelling artifacts from [82] .

A high signal is observed anterior to the medulla and at the junction of the medulla and upper cervical spinal cord on ASL maps (a). (b) Fast spin-echo FLAIR imaging and (c) T2-weighted imaging show no extra-axial lesion in this area, indicating that the signal originates within the CSF space.

2.5.2 Artifacts arising during transit of labelled spins

Loss of spin label

As labelled spins travel from the labelling plane to the imaging volume, they experience decay according to the blood T1 relaxation time, typically ranging from 1 to 2 seconds. Any factor that reduces the T1 relaxation time of blood can cause loss of the labelled spins. In clinical settings, the administration of gadolinium-based contrast agents is the most significant factor that reduces the T1 of intraarterial blood to approximately 100 ms. This rapid reduction causes labelled spins to revert to their equilibrium state before they reach the imaging volume, resulting in CBF maps devoid of signal, as illustrated in Figure 2.5.4. Gadolinium-based contrast agents significantly shorten the blood T1 relaxation time, causing labelled spins to rapidly return to their equilibrium magnetization state before they can reach the imaging volume, resulting in a loss of ASL signal. Consequently, it is crucial to conduct ASL

prior to the administration of contrast agents. If CBF maps display this specific signal absence, it is advisable to review the imaging protocol to confirm whether contrast was given before the ASL procedure, possibly during an earlier, separate session.

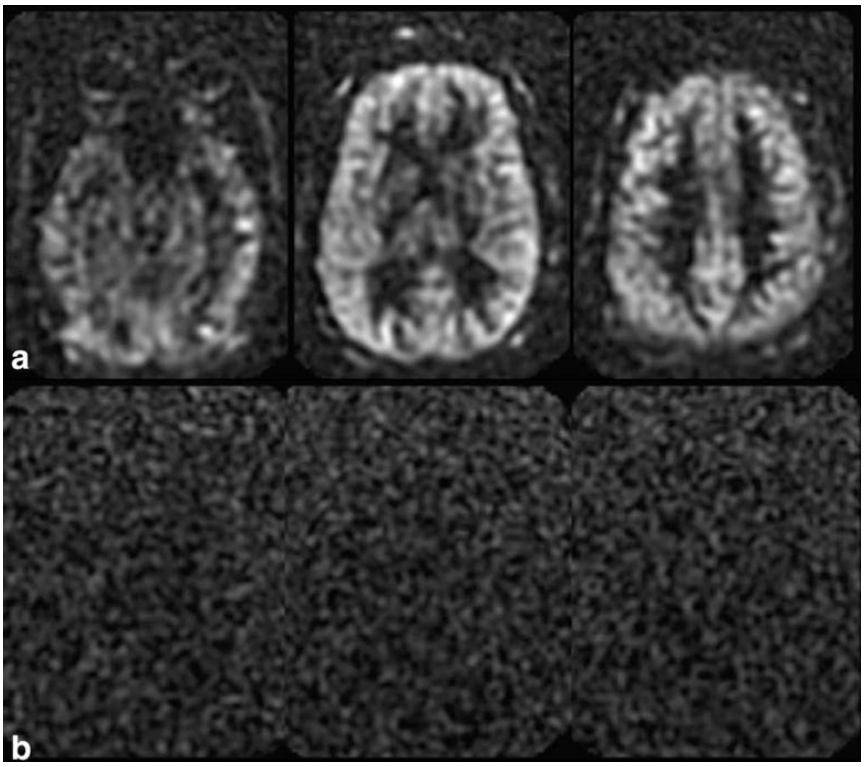


Figure 2.5.4 Signal loss due to the gadolinium-based contrast agents from [82] .

Before using contrast agent, the acquired ASL images (a)display a normal signal pattern. However, after contrast administration and ASL images (b) acquired, there is a complete absence of signal.

Arterial Transit artifact

The arterial transit time can be delayed due to conditions, such as reduced cardiac output, or arterial steno-occlusive disease, leading to artifacts for ASL. If the transit time of arterial blood surpasses the combined duration of the label time and PLD, imaging could occur prematurely, failing to capture parenchymal blood flow. In such cases, labelled spins remain within the intraarterial compartment and manifest as linear and serpiginous areas of high ASL signal in the CSF cisterns and cortical sulci, a phenomenon known as "arterial transit artifact" (ATA) (see Figure 2.5.5). Consequently, the parenchymal CBF distal to the ATA appears artifactually low because the labelled blood has not yet reached the parenchymal capillary bed for tissue exchange. This results in inaccurate measurements of parenchymal CBF,

particularly in conditions with extended transit times due to arterial narrowing or occlusion, posing a significant challenge to the use of ASL in determining tissue-at-risk using CBF thresholds in scenarios like acute stroke and chronic steno-occlusive disease. It is important to note that ATA may also occur in patients with normal cardiac output and unobstructed intracranial arteries if a short PLD (less than 1500 ms) is employed.

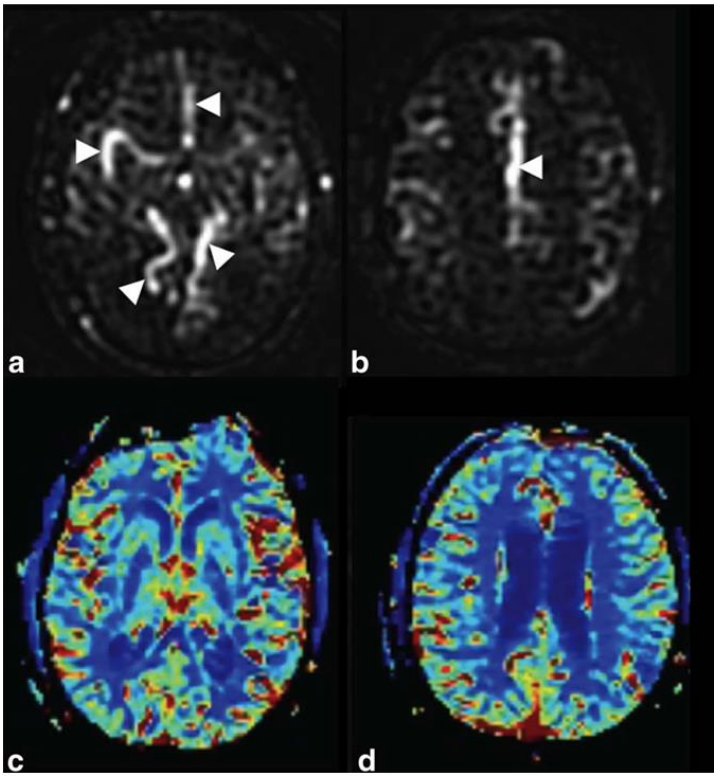


Figure 2.5.5 ATA and "ASL borderzone sign" in a patient with reduced cardiac output caused by cardiac failure from [82] .

ASL images (a, b) reveal the high signal intensity in the anterior, middle, and posterior cerebral arteries (arrowheads), indicative of the intraarterial retention of labelled spins, characteristic of ATA. Notably, there is a diminished signal in the brain parenchyma, especially within the parietal and occipital lobes, referred to as the "borderzone sign." DSC relative CBF maps (c, d) show no reduction in parieto-occipital blood flow, corroborating that the apparent perfusion deficit observed on ASL is an artifact resulting from the delayed transit of labelled spins into the parenchyma, rather than a true reduction in cerebral blood flow.

2.5.3 Artifacts arising during read-out

Several artifacts arise during the readout period, with their appearance influenced by factors such as motion and susceptibility gradients, which can vary to the readout technique employed.

Motion

Patient movement between control and tagging sequences in ASL often results in the appearance of a halo around the area of interest. In brain ASL, this artifact typically manifests as visualization of the scalp (see Figure 2.5.6). Additionally, movement can cause slice-to-slice variations in brightness. To reduce these effects, post-processing motion correction is frequently applied to align the paired images and minimize the impact of significant patient movement. Slices with substantial position changes can also be discarded with minimal effect on SNR, further enhancing the quality of the ASL data.

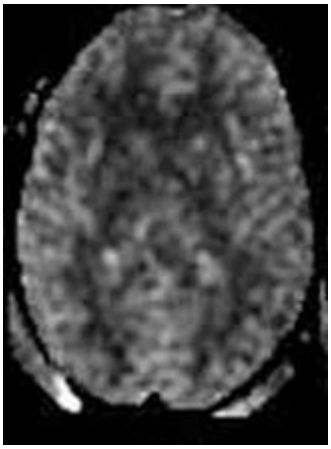


Figure 2.5.6 An example of axial CBF slice with motion artifact from [82] .
This was caused by varying contrast of ASL slices presented as bright halo at edge of images.

2.6 Quality Control

Quality assessment (QA) and quality control (QC) of MRI are crucial at various stages of the processing and analysis workflow to ensure the reliability of results. QA aims to avoid the occurrence of problems and to guarantee that the research workflow generates good quality data. Conversely, QC is to exclude poor-quality data from the dataset to ensure that such data will not be processed by subsequent analysis, thus potential bias in results can be prevented.

Most QC in MRI is conducted manually by trained technicians and radiologists (referred to “raters”), who strictly inspect the images to identify artifacts and other quality issues. This manual approach, while necessary, is heavily dependent on personal expertise, labor-intensive and subjective, leading to potential inconsistencies, inefficiencies and oversights in ensuring data quality. Thus, QC tools are developed to either reduce the time required for visual inspection or to

automatically evaluate the quality of MRI data. For this purpose, various image quality metrics (IQMs) are proposed to quantify MRI images from different perspectives, which serve as the primary input for QC tools. In the following section, prior research on MRI QC protocols and typical automated QC tools for various MRI modalities (MRIQC [83] for structural images, AFNI [84] for fMRI, and ExploreASL [85] as well as ASLPrep [86] for ASL MRI) will be introduced.

2.6.1 Quality Control Protocols

Many studies have proposed and applied various QC protocols to assist the screening and evaluate the image quality, aiming for automated QC for MRI data. For structural images, Woodard and Carley-Spencer [87] defined 239 no-reference IQMs based on Natural Scene Statistics or JPEG consortium to detect distortions and noise. Mortamet et al. [88] proposed QC protocols mainly focused on detecting specific artifacts (such as motion, blurring, and ghosting) in T1w images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [89] , by characterizing magnetization of the air background. This is based on the principle that most artifact signals propagate over the image and into the background. A study by Pizarro et al. [90] proposed three volumetric features (the normalized histogram, tissue-wise histogram, and the GM/WM mode ratio) and three artifact-specific features (eye motion spillover, head-motion spillover for ringing artifact, and wrap-around for aliasing artifact) for 1457 structural images. Furthermore, UK biobank study [171] combined imaging-derived phenotypes (IDPs) for T1w images, initially introduced as biomarkers to predict diseases, and additional QC-specific metrics (such as asymmetry between subcortical tissues, normalised intensity of each subcortical tissue, and pass/fail registration). Then an ensemble machine learning model combining three classifiers were trained with above QC metrics, and the voting system took the posteriori probabilities of these classifiers into account, following the “Minimum Probability” combination rule [91] [92] .

Recent studies also proposed QC protocols to incorporate fMRI images [93] [95] . In the study by [96] proposed Quality Assessment Protocol to measure the quality of both fMRI and structural images based on several quality metrics in the literature. The IQMs for structural images involve CNR, SNR, entropy focus criterion [97] ,

foreground-to-background energy ratio (FBER), voxel smoothness [98] , percentage of artifact voxels [88] . Regarding fMRI, spatial quality metrics were applied to the average image and temporal quality metrics were applied to fMRI data, these temporal quality metrics included the standardized root mean squared change in fMRI signal between volumes (DVARS) [99] , mean root mean square deviation (MeanFD) , the percentage of voxels with meanFD > 0.2 (Percent FD) [100] , the temporal mean of AFNI's 3dTqual metric (1 minus the Spearman correlation between each fMRI volume and the median volume) [95] and the average fraction of outliers found in each volume using AFNI's 3dTout command. Furthermore, the distributions of metrics were built, using the Autism Brain Imaging Data Exchange (ABIDE) [101] and the Consortium for Reliability and Reproducibility (CoRR) data [102] , to validate their reliability and reproducibility. A prior study [93] provided a visual assessment checklist using 129 resting-state fMRI data processed by Analysis of Functional Neuroimages (AFNI) to exclude outliers, which could be used to guide the training of new raters. In study [103] , resting-state and task fMRI data with various artifacts from public sources were used to conduct QC based on the visual assessment reports generated by MRIQC [83] and fMRIPrep [211] . Specifically, exclusion criteria were designed and applied for unprocessed fMRI data as well as T1w images based MRIQC visual report, and preprocessed fMRI data based on fMRIPrep visual report.

For ASL data, the consensus paper [1] outlined a manual approach for visual QC of ASL images, focusing on specific issues such as low labelling efficiency, global grey matter CBF values, and various artifacts. A later study [104] proposed a visual scoring system for ASL data involving two parts: contrast-based QC and artifact-based QC, which highlight the visual contrast between anatomical structures, and evaluate commonly occurred artifacts, respectively. Furthermore, apart from CBF maps, the scoring system also considers their ancillary parametric maps, if present, i.e. R1, aBV, and ATT maps. The detailed QC criteria are presented in Figure 2.6.1. The scoring system rates each item from the QC criteria according to its severity (higher scores mean better quality), with examples of artifacts at various degrees of severity provided as guides. This work provided relatively comprehensive QC protocols to measure the quality ASL data of various conditions, although it still requires manual inspection for individuals. Furthermore, ASL QC tools have been

developed, such as ExploreASL [85] and ASLPrep [86] . Both tools rely on their own processing pipelines, utilizing various image quality metrics derived from different stages of their processing to quantify the quality of ASL data.

Evaluation of CBF maps
Contrast component
Grey matter (0–2)
Grey/white matter differentiation (0–2)
Basal ganglia and thalami (0–2)
Subtotal (0–6)
Artifact component
Motion (0–2)
Signal drop (0–2)
Distortion (0–2)
Bright spots and areas (0–2)
Subtotal (0–8)
Grand total (0–14)
Evaluation of ancillary maps (QUASAR-specific)
Contrast
R1
Grey matter (0–2)
Grey/white matter differentiation (0–2)
Basal ganglia and thalami (0–2)
Subtotal (0–6)
aBV
Anterior cerebral arteries (0–2)
Middle cerebral arteries (0–2)
Posterior cerebral arteries (0–2)
Subtotal (0–6)
ATT
Anterior watershed area (2)
Posterior watershed area (2)
Deep watershed area (2)
Subtotal (0–6)
Grand total (0–18)

Figure 2.6.1 The QC criteria to evaluate CBF maps and ancillary images from [104] . R1: T1 relaxation rate; aBV: arterial blood volume; ATT: arterial transit time

2.6.2 Quality Control tools in MRI

2.6.2.1 MRIQC

MRIQC [83] , a quality control tool for structural MRI images, can detect outliers using a binary (accept/exclude) machine learning classifier trained with extracted IQMs. A minimal processing processing pipeline is provided to derive essential

corrected images required for IQM computation. Consequently, 64 metrics are extracted based on 14 IQMs of four families that measure noise, spatial distribution, particular artifacts, and other factors respectively, presented in Figure 2.6.2.

Furthermore, QC reports are provided by MRIQC: the individual anatomical report aims to facilitate the screening process and includes the image parameters, and visualization of cutting planes from structural images with mosaic views, contour of brain segmentation, and noise distribution on the air mask; the group report provides scatter plots that display the distribution of each IQM, aiding in the identification of outliers. These scatter plots are interactive, allowing users to open the corresponding individual report by clicking on the points within the plots.

Measures based on noise measurements	
CJV	The coefficient of joint variation of GM and WM was proposed as objective function by Ganzetti et al. [30] for the optimization of INU correction algorithms. Higher values are related to the presence of heavy head motion and large INU artifacts.
CNR	The contrast-to-noise ratio [31] is an extension of the SNR calculation to evaluate how separated the tissue distributions of GM and WM are. Higher values indicate better quality.
SNR	MRIQC includes the the signal-to-nose ratio calculation proposed by Dietrich et al. [32], using the air background as noise reference. Additionally, for images that have undergone some noise reduction processing, or the more complex noise realizations of current parallel acquisitions, a simplified calculation using the within tissue variance is also provided.
QI_2	The second quality index of [12] is a calculation of the goodness-of-fit of a χ^2 distribution on the air mask, once the artifactual intensities detected for computing the QI_1 index have been removed. The description of the QI_1 is found below.
Measures based on information theory	
EFC	The entropy-focus criterion [33] uses the Shannon entropy of voxel intensities as an indication of ghosting and blurring induced by head motion. Lower values are better.
FBER	The foreground-background energy ratio [14] is calculated as the mean energy of image values within the head relative the mean energy of image values in the air mask. Consequently, higher values are better.
Measures targeting specific artifacts	
INU	MRIQC measures the location and spread of the bias field extracted estimated by the inu correction. The smaller spreads located around 1.0 are better.
QI_1	The first quality index of [12] measures the amount of artifactual intensities in the air surrounding the head above the nasio-cerebellar axis. The smaller QI_1, the better.
WM2MAX	The white-matter to maximum intensity ratio is the median intensity within the WM mask over the 95% percentile of the full intensity distribution, that captures the existence of long tails due to hyper-intensity of the carotid vessels and fat. Values should be around the interval [0.6, 0.8].
Other measures	
FWHM	The full-width half-maximum [34] is an estimation of the blurriness of the image using AFNI's `3dFWHMx`. Smaller is better.
ICVs	Estimation of the icv of each tissue calculated on the FSL `FAST`'s segmentation. Normative values fall around 20%, 45% and 35% for cerebrospinal fluid (CSF), WM and GM, respectively.
rPVE	The residual partial volume effect feature is a tissue-wise sum of partial volumes that fall in the range [5%-95%] of the total volume of a pixel, computed on the partial volume maps generated by FSL `FAST`. Smaller residual partial volume effects (rPVEs) are better.
SSTATs	Several summary statistics (mean, standard deviation, percentiles 5% and 95%, and kurtosis) are computed within the following regions of interest: background, CSF, WM, and GM.
TPMs	Overlap of tissue probability maps estimated from the image and the corresponding maps from the ICBM nonlinear-asymmetric 2009c template [35].

Figure 2.6.2 IQMs from MRIQC [83] .

2.6.2.2 AFNI

AFNI (Analysis of Functional NeuroImages) [84] is a comprehensive software suite used for the analysis and visualization of fMRI data. AFNI provides a comprehensive set of quality control (QC) procedures for fMRI subjects [95] . The initial stage of QC, known as "getting to know your data" (GTKYD), occurs before any formal processing of datasets begins. Although not part of the inclusion/exclusion criteria, GTKYD presents and evaluates the consistency in acquisition parameters and data properties. Subsequently, systematic quantitative and qualitative stages are integrated directly within the afni_proc.py processing pipeline and QC HTML reports, labelled as APQUANT and APQUAL, respectively. For task-based fMRI, the STIM stage examines the stimulus event and timing information. Additionally, the graphical user interface (GUI) stage should be utilized for a subset of subjects in any study to thoroughly verify dataset properties and investigate any unknown features that may emerge during other QC stages.

2.6.2.3 ExploreASL

ExploreASL [85] is a software package for ASL image processing, that is developed using MATLAB based on Statistical Parameter Mapping (SPM). The quality control module is the final part of its processing pipeline, which considers both structural and ASL images. For structural images, QC parameters are adapted from Preprocessed Connectome Project Quality Assurance Protocol (QAP) [96] , presented in Table 2.6.1 from [85] . Specifically, a small WM region was used as the noise region instead of the background that out of the brain. On the other hand, for ASL images, QC parameters are adapted from SPM fMRI QC parameters [105] , presented in Table 2.6.2 from the supplementary material of [60]. Moreover, there are also QC parameters produced by comparing ASL images (after smoothing) with a group average (template) image (Table 2.6.3). The images with a large difference (> 2-3 SD) in QC parameters are suggested to be visually checked. These QC parameters as well as both T1w and ASL images (including intermediate and final images) were presented in a JSON file on individual level for visual QC.

Table 2.6.1 Structural QC parameters from ExploreASL [85] .

Structural QC parameters	Brief explanation
CNR_GM_WM_Ratio	abs(Mean GM - Mean WM)/(SD WMref)^2 (higher = better)
EFC_bits(bits)	Shannon entropy of voxel intensities proportional to maximum possible entropy for similarly sized image, indicating ghosting and head motion induced blurring (lower = better)
FBER_WMref_Ratio	(SD WholeBrain)2 / (SD WMref)2 (higher = better)
Mean_AI_Perc(%)	Mean of voxel-wise asymmetry index (AI), where AI = (L-R)/(0.5*[L+R]) (Kurth et al., 2015) (lower = better)
SD_AI_Perc(%)	Distribution of voxel-wise asymmetry index (AI), where AI = (L R)/(0.5*[L+R]) (lower = better)
SD_WMref(ml/100g/min)	SD in the WMref region
SNR_GM_Ratio	Mean GM / SD_WMref(higher = better)
WMref_vol_mL(mL)	Volume of noise reference region within the WM (WMref)
WMref_vol_Perc(%)	Percentage of WMref volume to the total WM volume

CNR=contrast-to-noise ratio, SD=standard deviation, EFC=entropy focus criterion, FBER=foreground to background energy ratio, AI=asymmetry index, Perc=Percentage, ref= reference, vol=Volume, L=left, R=right.

Table 2.6.2 ASL QC parameters from ExploreASL [85] .

ASL QC parameters	Brief explanation
tSNR_CSF_Ratio	Mean / temporal standard deviation (SD) within total CSF
tSNR_GM_Ratio	Mean / temporal SD within total GM

tSNR_GMWM_Ratio	Mean / temporal SD within the parenchyma (GM+WM)
tSNR_GMWM_WMref_Ratio	Mean within parenchyma / tSD within reference WM (WMref)
tSNR_Physio2Thermal_Ratio	tSNR_GMWM / tSNR_GMWM_WMref
tSNR_Slope_Corr	Slope of temporal SNR as function of GM partial volume
tSNR_WM_Ratio	Mean / temporal SD within total WM
tSNR_WMref_Ratio	Mean / temporal SD within WMref

Table 2.6.3 ASL difference QC parameters from ExploreASL [85] .

Difference QC parameters	Brief explanation
AI_Perc(%)	Mean voxel-wise asymmetry index (AI), computed as (L-R)/(0.5*[L+R]) (lower = better)
Mean_SSIM_Perc(%)	Mean voxel-wise structural similarity index (SSIM) (higher = better)
nRMSE_Perc(%)	Normalized Root-mean-square error (nRMSE) with ASL template (lower = better)
PeakSNR_Ratio	Assuming an individual ASL image can be interpreted as a noisy version of an ASL template, this parameter is the dynamic range compared to the mean squared difference of the individual and template ASL images. Dynamic range is calculated here as MaxIntensity-MinIntensity (higher = better)
RMSE_Perc(%)	Root-mean-square error (RMSE) with ASL template (lower = better)

2.6.2.4 ASLprep

ASLPrep [86] is a software workflow for ASL MRI data processing that is developed using python package Nipype. In ASLPrep, both importing metadata and storing processed data conform to the Brain Imaging Data Structure (BIDS) specification [106] for ASL.

In QC workflow, ASLPrep measures quality of registrations of ASL to T1w image as well as T1w image to the standard template for each step during registration, theses metrics include the mask overlaps, spatial correlation, Dice coefficient and Jaccard index. Furthermore, ASLPrep also provides quality measures for ASL timeseries which are the mean frame-wise displacement, the root mean square variance of temporal derivative of CBF time courses, CBF ratio between GM and WM, and the CBF quality evaluation index (QEI). QEI measures the quality of CBF map using its similarity with structural tissues, CBF spatial variabilities in GM and WM, and the percentage of negative CBF within the GM mask.

Furthermore, a descriptive HTML report is generated on individual level. Specially, the report includes the visualization of data before and after key processing steps (normalization, registration, distortion correction) as well as CBF maps (SCORE-corrected CBF [107] , SCRUB-corrected CBF[108] , BASIL [109] CBF, and PVEc CBF), aforementioned quality control metrics, and the description of all methods/tools used for preprocessing with citations.

2.7 Automated Quality Control Techniques

Conventional methods are widely used for QC in MRI data and play a crucial role in ensuring data integrity. One common approach is the application of z-score thresholds to identify outliers by measuring the deviation of individual data points from the mean [110] . This statistical method effectively highlights anomalies that could indicate potential issues in the MRI data. Another traditional technique involves the use of boxplots, which visually represent the distribution of data and identify outliers [111] . By depicting the spread and central tendency of the data, boxplots offer a straightforward means of detecting anomalies that might compromise image quality. Additionally, methods based on Principal Components

Analysis (PCA) are employed to facilitate anomaly detection in MRI [112] ,[113] . PCA reduces the dimensionality of the data, helping to identify patterns and variances that may indicate anomalies. This approach is particularly useful in isolating noise and artifacts that can adversely affect MRI image quality.

While these conventional methods are valuable, they often require significant manual intervention and may not always capture complex patterns in the data. This limitation has motivated the use of machine learning and deep learning techniques [114] [117] for quality control in MRI. Machine learning is particularly effective in selecting the most useful features, optimizing the process by automatically identifying and prioritizing relevant metrics that contribute to data quality and making it more efficient and less prone to human error. In addition, deep learning takes this a step further by automatically learning patterns directly from the image data. These models can identify subtle anomalies that traditional methods might miss, thanks to their ability to process large amounts of data and discern intricate relationships within it. This makes deep learning particularly powerful in the context of MRI, where detecting slight irregularities can be crucial.

In deep learning, the process of QC in MRI images is often referred to as anomaly detection or out-of-distribution (OOD) detection. Anomaly detection in MRI images is typically conducted using convolutional neural network [118] (CNN)-based feature extractors followed by classifiers. CNNs excel at identifying spatial hierarchies in images, making them particularly suited for detecting irregularities in MRI scans. Additionally, deviation-based models such as variational autoencoders (VAEs) [119] and generative adversarial networks (GANs) [120] are also employed. These models can generate synthetic data that closely resembles the training data, enabling them to identify deviations that indicate anomalies. By leveraging these advanced techniques, MRI QC can become more automated, accurate, and capable of handling the complexities inherent in medical imaging.

2.7.1 Machine Learning

Machine learning is a branch of artificial intelligence that involves creating algorithms and statistical models that allow computers to learn from data and make predictions or decisions. It is generally divided into two main categories: supervised

learning, where models are trained on labelled data to predict specific outcomes, and unsupervised learning, which focuses on discovering hidden patterns or structures in unlabelled data. QC of MRI data is to accept normal data and exclude outliers, and thus it can be regarded as a binary classification problem in machine learning.

A machine learning (supervised) task commonly involves several key steps: Initially, data collection is undertaken to obtain the dataset comprising input features and their corresponding labels. In MRI studies, IQMs are extensively utilized as features for machine learning models to assess and ensure the quality of MRI data. The corresponding labels for these machine learning tasks are typically provided by expert raters. Subsequently, data preprocessing is conducted to clean and prepare the data, including handling missing values, normalizing numerical features, and encoding categorical variables. Feature selection or engineering follows, aiming to identify and create the most relevant features that will improve model performance. The next step involves splitting the dataset into training and validation sets, which allows for both model training and performance evaluation. The chosen classification algorithm is then trained on the training data, learning to map input features to the corresponding labels. Hyperparameter tuning is performed to optimize the parameters of model, often using techniques such as grid search or cross-validation. After training, the performance is evaluated using the validation set, employing metrics such as accuracy, precision, recall, and the F1 score. The final step is model testing, where model generalization ability is assessed on an independent test set to ensure its reliability on unseen data.

Notably, prevalent challenge in this domain is the inherently imbalanced dataset, where the majority of images may fall into the category of acceptable quality, while a smaller proportion exhibit significant artifacts or quality issues. This imbalance can lead to biased models that perform well on the majority class but poorly on the minority class, thereby necessitating careful selection of the model to address this challenge. Support Vector Machines (SVMs) [121] are commonly employed in these contexts due to their effectiveness and robustness in handling imbalanced data. Further details of SVMs are given in the following section.

Support Vector Machine

Support vector machines (SVMs) [121] are widely used for classification tasks, which is achieved by identifying the optimal hyperplane that separates two classes, aiming to maximize the margin between the nearest data points from each (see Figure 2.7.1). Although many potential hyperplanes could separate the classes, SVM seeks the one that provides the largest margin, thus ensuring the best decision boundary. This maximization of the margin helps the model generalize well to new data, to achieve accurate classification predictions. The lines adjacent to the optimal hyperplane are called support vectors, as they pass through the data points that are crucial for defining the margin. This allows the model to maintain sensitivity to the minority class (outliers) [122] .

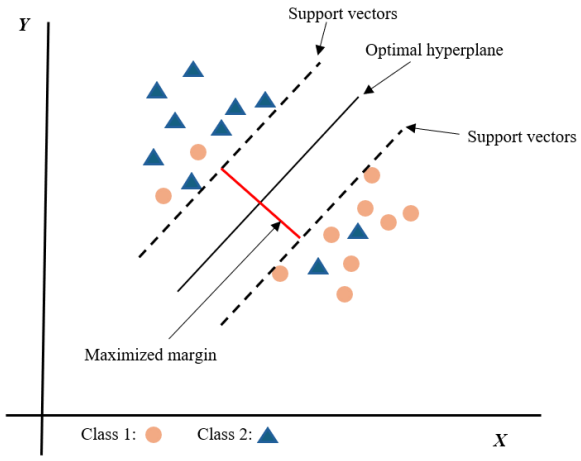


Figure 2.7.1 SVM classifies data by maximizing the margin between the two classes.

The middle line between the support vectors is the optimal hyperplane and the distance between the supports is denoted as maximized margin.

In the binary classification, we suppose there are two classes with labels as +1 and -1. **x** denotes the input features and **y** is the corresponding labels. The hyperplane can be expressed as:

$$w^T x + b = 0 \qquad \text{Equation 2.7.1}$$

The vector w denotes the normal vector that is perpendicular to the hyperplane and the parameter b denotes the distance along the vector w to the hyperplane, which can be represented as:

$$d_i = \frac{w^T x_i + b}{||w||} \qquad \text{Equation 2.7.2}$$

The margin can be further denoted as:

$$\gamma_i = \frac{y_i(w^T x_i + b)}{||w||} \qquad \text{Equation 2.7.3}$$

Where γ_i is the margin and y_i is the label of an example. In linear hard margin SVM classifier, the prediction of an example can be denoted as:

$$\hat{y} = \begin{cases} 1 : w^T x_i + b \geq 0 \\ -1 : w^T x_i + b < 0 \end{cases} \qquad \text{Equation 2.7.4}$$

The goal of SVM is to maximize the margin between the two classes: in Equation 2.7.5, $min_i|w^T x_i + b|$ can be regarded as 1 by changing the scale of vector w, and thus $max_{w,b} \frac{1}{||w||}$ can be converted to $min||w||$, and end up with $min\frac{1}{2}||w||^2$, which can be solved as a convex quadratic programming problem.

$$max_{w,b}\, min_i \frac{1}{||w||}|w^T x_i + b| = min_i \frac{1}{2} w^T w = \; min_i \frac{1}{2}||w||^2,$$

$$subject\ to\ y_i(w^T x_i + b) > 0\ for\ i = 1,2,3 \ldots, m \qquad \text{Equation 2.7.5}$$

In many real-world scenarios, data is often not linearly separable in its original feature space, presenting a significant challenge for traditional classification methods. For instance, extreme high or low values in quality metrics may indicate outliers that lie on opposite ends of the data distribution, complicating the separation of classes. The kernel trick in Support Vector Machines (SVMs) effectively addresses this challenge by implicitly mapping input data into a higher-dimensional feature space, where a linear separator can be more easily identified. Instead of explicitly computing the coordinates of the data in this higher-dimensional space—a task that can be computationally prohibitive—the kernel trick allows SVMs to operate in this transformed space by evaluating a kernel function. This function effectively computes the inner product of the data points in the high-dimensional space, enabling the SVM to find an optimal hyperplane without ever needing to directly perform the transformation. Several types of kernel functions are commonly used, including the linear, polynomial, radial basis function (RBF), and sigmoid kernels. Each of these kernels introduces a different notion of similarity in the transformed space, allowing the SVM to adapt to various types of data distributions.

By leveraging the kernel trick, SVMs can efficiently handle complex classification tasks that involve non-linear decision boundaries. This capability makes SVMs a powerful tool in machine learning, particularly in domains where the relationships between features are not straightforward or easily captured by linear models. The kernel trick, therefore, not only extends the applicability of SVMs but also enhances their performance on a broad range of classification problems.

2.7.2 Deep Learning

Deep learning is a subset of machine learning that focuses on using artificial neural networks with many layers, also known as deep neural networks, to model complex patterns in data. It is part of a broader family of machine learning methods and is particularly effective in tasks that involve large amounts of data and require learning hierarchical representations, such as image recognition, natural language processing, and speech recognition. Recently, deep learning has emerged as a powerful approach for QC in medical images [114] [115] . Unlike classical machine learning techniques that often rely on predefined IQMs, deep learning can inherently capture complex image-level features, without the need for such predefined QC criteria.

The training process of neural networks is analogous to machine learning in that it involves iteratively adjusting the model parameters to minimize the error in predictions. This process begins with the collection and preprocessing of a dataset, which is then used to train the network. The data are fed into the neural network in a process called forward propagation, where it passes through multiple layers of neurons. The accuracy of these predictions is evaluated using a loss function, which quantifies the difference between the predicted outputs and the actual targets. Through backward propagation, the gradients of the loss function with respect to the weights of neurons in each layer are computed sequentially from back to front, allowing for the adjustment of weights in the direction that reduces the loss. This optimization is typically performed using algorithms like Stochastic Gradient Descent [123] or Adam [124] . Techniques such as data augmentation, regularization, and cross-validation in this training phase can enhance the generalization of models for unseen data and does not overfit. Techniques like early

stopping, learning rate scheduling, and hyperparameter tuning can further refine the training process, helping to achieve optimal performance.

This section provides an overview of these fundamental deep learning concepts, which serve as the theoretical basis for QC using deep learning of ASL MRI in Chapter 5.

2.7.2.1 Perceptron

A perceptron [125] is a basic building block in deep learning, representing the simplest form of an artificial neuron or neural network unit. It serves as the foundation for more complex neural network architectures, such as multi-layer perceptrons (MLPs) [125] and convolutional neural networks (CNNs) [118] .

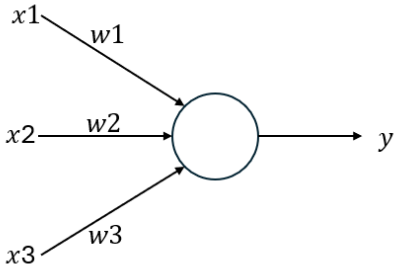


Figure 2.7.2 Perceptron model.

***x** is the input vector, and **w** is the corresponding weight vector. y is the output.*

A perceptron is a mathematical model of a neuron, designed to simulate the behaviour of biological neurons in the brain, presented in Figure 2.7.2. It takes a set of input values, applies a linear transformation (through a set of weights), and then uses an activation function to produce typically an output of 1 (active) or 0 (inactive).

$$y = \begin{cases} 0 & (\sum_j w_j x_j \leq \theta) \\ 1 & (\sum_j w_j x_j > \theta) \end{cases} \qquad \text{Equation 2.7.6}$$

The perceptron calculates a weighted sum of its inputs, which is then passed through an activation function to generate the final output. Perceptrons laid the groundwork for the development of deep learning, demonstrating the potential for neural networks to learn from data and perform complex tasks.

2.7.2.2 Multi-layer Perceptrons

This simple structure of perceptron allows linear classification, but they struggle with non-linear problems. To address this limitation, MLPs (depicted in Figure 2.7.3) were developed, incorporating multiple layers of neurons. The MLPs include one hidden layer, enabling complex non-linear modelling. Despite their simplicity, MLPs serve as the foundation for modern deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which build on these concepts to tackle a wide range of complex tasks in machine learning and artificial intelligence.

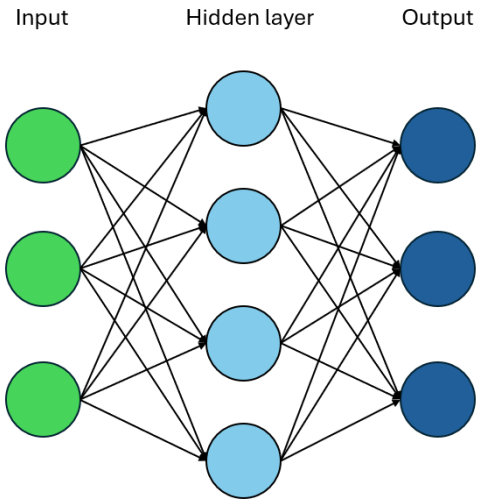


Figure 2.7.3 Multi-Layer Perceptrons including input layer, hidden layer and output layer from left to right.

2.7.2.3 Convolutional Neural Network

In a simple neural network, each node in one layer is connected to every node in the next layer, and there is typically only a single hidden layer. In contrast, deep learning networks have multiple hidden layers, making them "deep." These additional layers allow the network to learn more abstract and complex representations of the input data. Although basic neural networks, like feedforward networks, can process images as input, they are highly inefficient due to the full connection between all neurons across layers. Deep learning architectures, such as Convolutional Neural Networks (CNNs) [118] , overcome this limitation by preserving spatial information and establishing sparse connections between neurons in different layers, making them more efficient and effective for image processing tasks (such as Figure 2.7.4).

A characteristic of CNNs is use of convolutional layers, where filters (also known as kernels) are applied to input data to extract relevant features or preserve the relationship of local areas, such as edges, textures, or patterns. Each convolution kernel is a feature extractor, and only one feature is extracted. These filters slide across the input, performing element-wise multiplications and summations, producing feature maps that highlight important aspects of the data. CNNs also incorporate pooling layers, which reduce the dimensionality of the feature maps by down-sampling, thus lowering computational complexity while preserving the most critical information. The combination of features is the key to determine what the original image is, and the absolute position of features is less important while the relative positioning between different features plays a more significant role. Using the relative position can be more robust to some translation and rotation, control overfitting, increase the generalization ability of the model, and greatly reduce the model parameters. Another common feature of CNN is the use of non-linear activation functions like ReLU (Rectified Linear Unit) [118] , which introduce non-linearity to the model, allowing it to learn complex patterns. CNNs often conclude with fully connected layers, facilitating final decision-making or classification tasks [118] . Together, these features make CNNs highly effective for tasks involving image processing and computer vision, enabling them to excel in applications like image classification, object detection, and semantic segmentation.

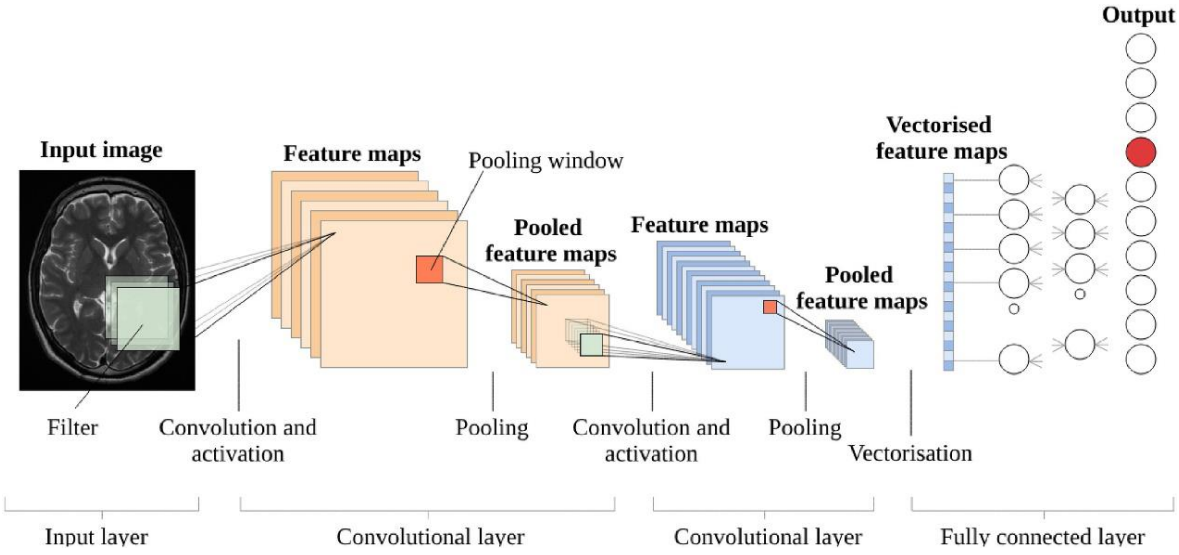


Figure 2.7.4 The conventional blocks of a typical CNN for medical image process from [127] .

2.7.2.4 U-Net

The U-net architecture [128] is widely recognized for its efficacy in medical image segmentation (Figure 2.7.5). This neural network structure features a symmetric, U-shaped design that facilitates the efficient processing of images through a combination of contraction and expansion paths. The contraction path of the U-net is akin to a traditional convolutional network, comprising repeated application of convolution and pooling layers. This path captures context by downsampling the input image, effectively learning hierarchical features. Conversely, the expansion path consists of upsampling operations that increase the resolution of the output. This path merges high-level feature representations with corresponding features from the contraction path through skip connections. These connections preserve spatial information that is crucial for accurate segmentation.

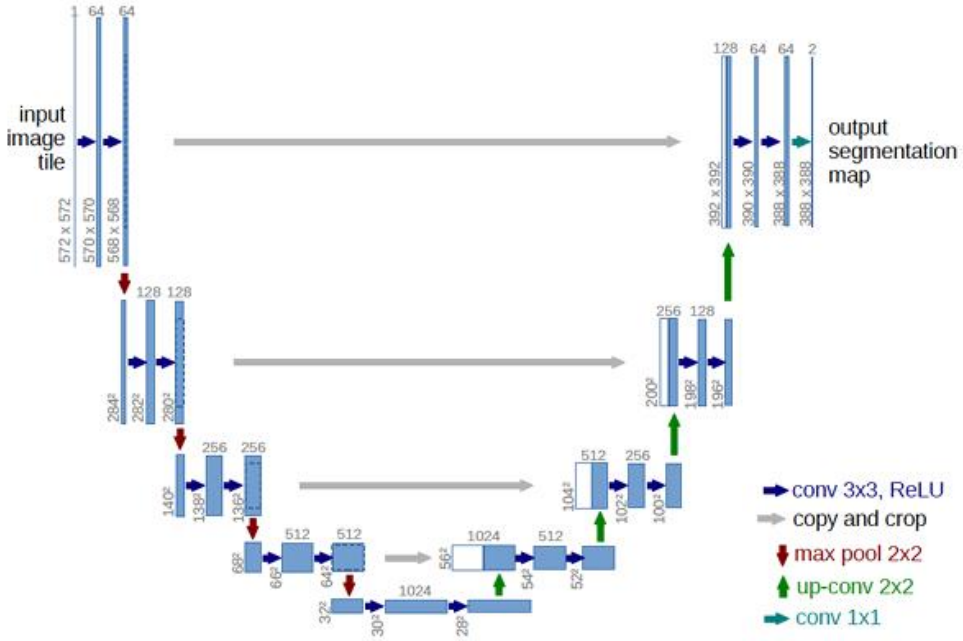


Figure 2.7.5 U-net architecture from [128] .

The network consists of an encoder-decoder structure with symmetrical skip connections that help preserve spatial information. The encoder (contracting path) extracts hierarchical features through successive convolutional and pooling layers, reducing spatial dimensions while increasing feature complexity. The decoder (expanding path) restores spatial resolution using transposed convolutions and combines low-level features from the encoder via skip connections to refine the output. The final layer applies a convolution operation to generate the segmented output, making U-Net highly effective for medical image analysis tasks.

2.7.3 Anomaly Detection

Quality control, often referred to as anomaly detection in the context of deep learning, presents unique challenges due to the difficulty of obtaining labelled data for every possible anomaly. While supervised learning algorithms, particularly those implemented through CNNs, have seen widespread use in industries such as manufacturing, surface inspection, and surveillance [129] [131] , however, their applicability in medical imaging is limited. The primary obstacle is the impracticality of acquiring comprehensive labelled datasets for all potential anomalies in MRI scans.

Consequently, supervised learning methods [132] ,[133] are not ideally suited for anomaly detection in medical imaging. Instead, semi-supervised learning approaches [133] [134] have emerged as a more viable solution. These methods focus on identifying deviations from established norms, rather than relying on extensive labelled data. The initial step in this process involves developing a robust baseline of normal MRI scans, which serves as a reference point for detecting anomalies. This baseline is created by analysing a large dataset of MRI images that have been verified to be free of defects. Deep learning algorithms are employed to learn the typical patterns and features of these high-quality MRI images, establishing a standard against which new, unseen images can be compared. Once this baseline is established, the anomaly detection system can evaluate unseen MRI data by identifying deviations from the norm. Any irregularities that suggest potential anomalies are flagged by comparing the new images against the established baseline. This semi-supervised approach thus enables the detection of anomalies without the need for extensive labelled datasets.

In addition to semi-supervised methods [135] [137] , unsupervised anomaly detection algorithms are also employed. These algorithms analyse both normal and anomalous data without relying on labels, identifying anomalies based on intrinsic properties of the datasets, such as distances or densities. By examining the structure and distribution of the data, unsupervised methods can detect outliers that do not conform to established patterns.

The combination of semi-supervised and unsupervised approaches provides a comprehensive strategy for anomaly detection in MRI. By leveraging both defined

norms and intrinsic data properties, these methods enhance the reliability and accuracy of detecting anomalies in medical imaging.

2.8 Conclusion

This chapter outlined the fundamental concepts of MRI, with a specific focus on ASL MRI that stands out as a non-invasive technique that enables the quantitative measurement of perfusion. The ASL consensus paper [1] on the clinical implementation of ASL MRI marked a pivotal milestone in 2015, promoting its broader acceptance in clinical practice and establishing standardized protocols for both research and routine clinical applications [5] [34] [54] . Since then, various advancements have been made to ASL to enhance image quality, improve the accuracy of CBF quantification, and provide measurements of additional physiological parameters [27] . However, ASL still faces several challenges. This thesis focuses on two main aspects that complicate ASL MRI: partial volume effect correction and automated quality control methods.

PVEc methods, such as linear regression [7] and spatially regularized variational Bayes approaches [4] , are well-established in the literature. However, the impact of PVEc on ASL results is still debated, potentially due to factors like sample size or variability in protocols. One key limitation is the relatively small participant numbers [177] ,[181] [186] , which restricts the generalizability of findings and complicates the assessment of effects of PVEc.

Another challenge in ASL MRI is automated QC. This chapter discussed some common artifacts that can arise during ASL acquisitions and highlighted the limitations of traditional QC methods. Automated QC, especially with the integration of advancing deep learning techniques, shows promise in detecting artifacts and evaluating image quality for subsequent analysis.

To address these challenges, Chapter 3 applied PVEc to a relatively large ASL data, and investigated age-related changes in perfusion with and without PVEc to account for anatomical changes, such as brain atrophy, which can lead to cortical thinning and exacerbate GM voxel contamination by other tissue types. Additionally, a surface-based pipeline for ASL pre-processing and analysis was developed, alongside

the conventional volumetric pipeline, to provide surface-based representations of the results. Chapter 4 explores image quality metrics specific to ASL data to aid in QC, with the goal of standardizing ASL QC protocols, a gap that currently exists in the literature. These IQMs were incorporated into QC reports for quick visual inspection. Moreover, these metrics were individually evaluated and also used as features for machine learning tasks. In Chapter 5, deep learning models were built to automatically learn features directly from the images, eliminating the need for predefined IQMs, thereby enabling fully automated QC for ASL MRI.

3 Regional Changes in Cerebral Perfusion with Age When Accounting for Changes in Grey Matter Volume[1]

3.1 Introduction

Perfusion is the process of delivering blood to the capillaries within tissues, essential for maintaining tissue and organ health by supplying oxygen and nutrients while removing metabolic waste [138] [141] ,[142] . Cerebral blood flow (CBF), a quantitative measure of perfusion, represents the rate at which arterial blood flows through capillaries and is closely tied to normal metabolic function [146] -[148] . Arterial Spin Labelling (ASL) MRI is the only non-invasive technique that allows imaging of CBF in the brain, using blood water as an endogenous tracer [34] . ASL has become a widely used method for assessing CBF in various neurological conditions, as well as for studying the effects of age and sex on cerebral perfusion. While there is general agreement that CBF decreases with age, research findings on the influence of age and sex on CBF in specific brain regions have been inconsistent. Some studies indicate a decline in cortical CBF with age, while others report an increase or no significant correlation [150] . These conflicting results are likely due to limited cohort sizes and variability in the analysis methods employed.

Like other perfusion techniques, ASL has a relatively low spatial resolution typically with voxel size around 3–5 mm. This resolution is lower than the anatomical variations within the tissues it aims to image. This can significantly impact CBF measurements affected by partial volume effects (PVE) since a voxel is likely to contain more than one type of tissues. Although this might not be evidently problematic when visually examining areas of hyper- or hypo-perfusion, such as brain tumors or stroke, PVE can confuse the quantitative measurements of CBF, especially when comparing individuals or detecting subtle perfusion changes. This is

[1] This work has been submitted as a paper to Magnetic Resonance in Medicine [242] .

particularly relevant in cases of neurodegenerative conditions, such as Parkinson's disease [6] , or certain stroke effects. In these scenarios, actual changes in blood flow might be misinterpreted for, or masked by CBF differences or changes between corresponding voxels in the same anatomical voxels. It is increasingly recognized that correcting PVE is crucial to quantify GM CBF independently of any confounding effects from partial volumes of WM or CSF [4] [175] . Therefore, various partial volume effects correction (PVEc) methods have been developed that use voxel-wise estimates of PVs to identify the signal from each tissue. Examples include the Muller-Gartner method for PET [153] , and linear regression [7] or spatially regularized variational Bayes methods for ASL [8] .

Traditional methods for obtaining partial volume estimates rely on volumetric segmentation, such methods may not provide precise results for intricate structures like the cortex, with its efficacy being contingent on the accuracy of the segmentation method employed. Given the complexity of shapes such as the thin, intricately folded cerebral cortex, surface-based segmentation has become increasingly popular, especially with tools like FreeSurfer [154] . This method offers significant advantages. First, it allows for more continuous representation because it places surface vertices with precision finer than a voxel, in contrast to the inherently discrete nature of volumetric segmentation. Second, it enables the application of anatomically accurate constraints that vary directionally, such as ensuring tissue homogeneity along a surface while allowing for heterogeneity perpendicular to it. While there are surface-based partial volume estimation tools in existing studies, previous attempts have typically been tailored for a particular imaging modality, e.g., the Human Connectome Project's (HCP) fMRI Surface pipeline for BOLD using ribbon-constrained method [155] and PETSurfer [156] , a variant of FreeSurfer, for PET.

Current brain research using ASL predominantly relies on region of interest (ROI) analysis or volumetric (i.e., voxel-based) analysis. In the ROI approach, researchers define anatomical or functional brain regions and compute a quantity of interest within these regions. This method, however, is limited by the prior hypotheses to determine the expected ROI. On the other hand, the volumetric method has an advantage in identifying unanticipated or non-hypothesized areas of abnormal activity in the brain [159] . Surface-based analysis has become increasingly popular

in neuroimaging because of its effectiveness in representing the cortex [159] . The cortex is intricately connected to various cognitive functions, including perception, attention, and language. The traditional volumetric analysis defines a neighbourhood in a three-dimensional space, including every voxel within a designated sphere around a central point, without considering if these voxels share the same tissue type to the central point. Alternatively, surface-based analysis confines the neighborhood specially to the cortical surface, or "cortical ribbon," and calculates distances exclusively along this ribbon. This method prevents the mixing of neighboring WM, CSF, and subcortical GM from cortical GM and avoids confounding between cortical areas that are close in Euclidean space but distant along the cortical surface.

Some studies have used cortical surface-based analysis in perfusion to investigate regional CBF changing with aging [157] -[160] . However, relatively few studies applied this technique to ASL data specifically [150] ,[161] . A prior study demonstrated that performing cortical surface-based analysis on ASL is technically feasible and produces high-quality images. This approach has the potential to significantly improve the detection of focal perfusion changes, especially in neurodegenerative diseases, within a clinical setting [162] . Despite these promising findings, surface-based analysis has not yet been widely integrated into ASL research. The existing gap in the literature underscores the importance of adopting surface-based analysis in ASL studies. By leveraging the detailed cortical geometry inherent to surface-based approaches, this method holds promise for delivering enhanced precision in investigating disease mechanisms, tracking progression, and treatment effects.

A previous study focusing on the elderly cohort from TILDA ASL data [163] investigated age-related perfusion changes and reported on mean global brain perfusion without PVEc. The objective of current study was to perform a subsequent exploratory investigation of the influence of aging on regional brain perfusion with PVEc in the same elderly dataset, potentially accounting for the influence of changes in anatomy. This study proposed a surface-based pipeline along with the conventional volumetric pipeline for ASL data (see Figure 3.2.1), aiming to demonstrate the processing and results with the two different representations for further use and interpretation by the community.

3.2 Methods

3.2.1 Dataset

This study used a subset of participants from The Irish Longitudinal Study on Ageing (TILDA), a prospective cohort study that collects health, economic, and social aspects of Irish adults [166] ,[167] . The characteristics of the dataset, including age, sex, education, and health conditions, were detailed in a prior study using the same dataset [163] . The exclusion criteria were also detailed in the prior study [163] . Briefly, participants were excluded if they presented any contraindications to MRI, a history of stroke or head injury, or if their images contained artifacts. To ensure the reliability of the analysis, subjects outside the specified age range of 54 to 84 years were excluded. Furthermore, participants with extreme mean GM CBF values in non-PVEc data (<10 or >100 ml/100g/min) were also excluded. This exclusion criterion led to a discrepancy in the total number of participants compared to the prior study [163] due to different calibration methods resulting in differences of perfusion estimates. Consequently, the final sample comprised 423 healthy participants (215 females and 208 males) aged 54 to 84 years, all of whom had available T1-weighted and pCASL data.

3.2.2 MRI data acquisition

All MRI data (T1-weighted and pseudo-continuo ASL (pCASL) sequences) were acquired on a 3T scanner, using a 32-channel head coil. T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) anatomical images were acquired over 5 min 24 s with the following scan parameters: field of view (FOV) = 240 × 240 × 162 mm^3, matrix = 288 × 288 × 180, repetition time (TR) = 6.7 ms, echo time (TE) = 3.1 ms, flip angle = 8°, and SENSE = 2. The pCASL acquisition parameters using 2D multislice single-shot EPI were as follows: 30 interleaved pairs of images acquired alternating with and without arterial spin labelling, FOV = 240 × 240 mm^2, matrix=80×80, TR=4000ms, TE=9ms, FA=90°, SENSE=2.5, and scan duration = 4 min 16 s. 13 slices (8 mm thick, 1 mm gap) were acquired sequentially in a caudocranial direction. A labelling duration of 1800 ms and a post-label delay of 1800 ms were used. Calibration scans measuring the equilibrium magnetisation (M_0)

were also acquired using the same geometry as the pCASL sequence, with TR = 10,000 ms, TE = 9 ms, and scan duration = 20 s. B_0 field maps were measured using a two-echo 2D gradient echo sequence with the same in-plane resolution as the pCASL scans and the following acquisition parameters: TR=455ms, TE1 /TE2 =1.69/7.0ms, FA=90°, and scan duration = 39 s.

3.2.3 Data processing

The overall ASL data processing pipeline used in this study was shown in Figure 3.2.1.

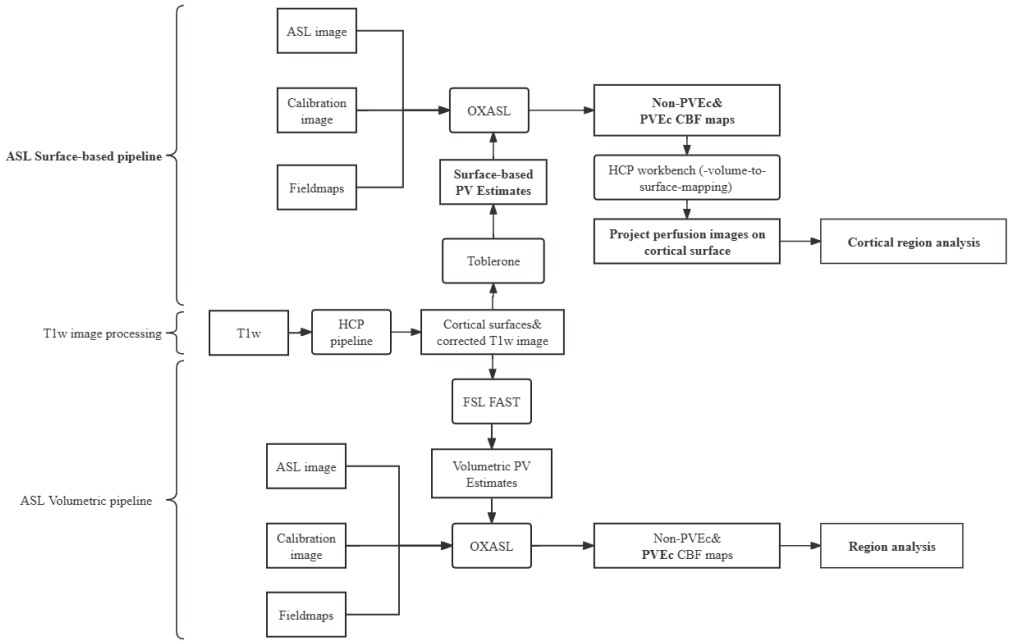


Figure 3.2.1 ASL pipelines for this study.

Top: the proposed ASL surface-based pipeline; Middle: T1w pre-processing pipeline from Human Connectome Project [231] ; Bottom: the conventional ASL volumetric pipeline [109] with PVEc.

3.2.3.1 Structural image processing

T1-weighted images were processed using the HCP minimal processing pipelines [155] ,[168] , which consist of three distinct stages: (1) the PreFreeSurfer pipeline, which corrects for gradient distortions and bias fields (B_1 inhomogeneities), performs brain extraction, and aligns the images to the MNI152 template space [70] ; (2) the FreeSurfer pipeline, which segments structural volumes according to a specified

parcellation, reconstructs white and pial cortical surfaces, and registers these surfaces to the FreeSurfer surface atlas; and (3) the PostFreeSurfer pipeline, which generates the final NIFTI files and GIFTI surface files registered to the Conte69 surface atlas [155] [169] .

3.2.3.2 Partial volume estimation

To obtain partial volume estimates, the volumetric pipeline used partial volume estimates (GM/WM/CSF) calculated by FSL FAST (https://fsl.fmrib.ox.ac.uk/fsl/oldwiki/FAST.html) [164] with the pre-processed T1w image (see Figure 3.2.2). The surface-based pipeline utilized Toblerone (https://toblerone.readthedocs.io/) [165] to calculate GM/WM partial volumes, leveraging the structural information provided by Freesurfer (https://surfer.nmr.mgh.harvard.edu/) [154] (see Figure 3.2.2). This approach allowed us to estimate partial volumes within the cortex and all structures identified by Freesurfer in the structural space.

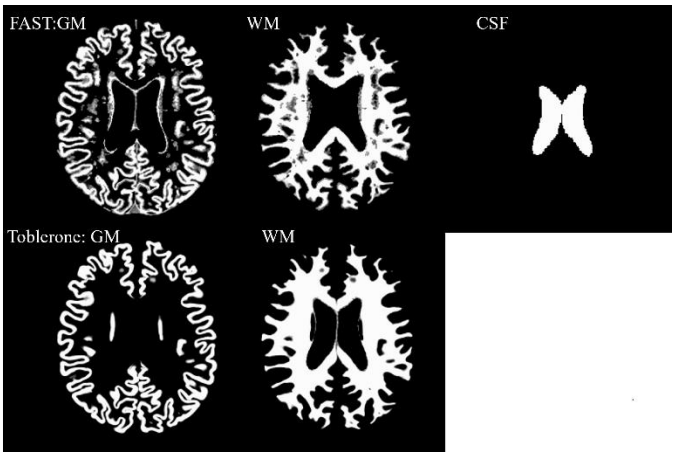


Figure 3.2.2 Partial volume estimates in the from FSL FAST and Toblerone.

FSL FAST PVs (top from left to right):GM, WM, and CSF; Toblerone PVs (bottom from left to right): GM, WM.

3.2.3.3 ASL image processing

With corrected T1w image and its partial volume estimates, pCASL images were processed using OXASL (https://github.com/physimals/oxasl) a pipeline for performing Bayesian analysis of Arterial Spin Labelling MRI data, an updated version of the BASIL toolbox found in FSL [109] , giving access to more advanced

regions and surface analysis options. Within the OXASL pipeline, motion correction was firstly calculated for ASL data using FSL MCFLIRT, with the calibration image as reference. Then the perfusion-weighted image derived by averaging the ASL difference images was registered to the corrected T1w image using FSL FLIRT (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT) to obtain the transformation between native space and structural space. B_0 maps were utilized to calculate distortion corrections for the ASL data, compensating for any spatially nonlinear image distortions caused by B_0 inhomogeneities in these EPI data. At the end of pre-processing, these corrections were applied to ASL data and calibration data. Quantification of the ASL data was conducted in native space using Bayesian inference, following the standard well-mixed single compartment kinetic model without dispersion of the bolus of labelled blood water [170] . Voxel-wise calibration was used to calculate absolute perfusion values (CBF in ml/100g/min) using the calibration image (M0) to calculate the scaling factor on a voxel-by-voxel basis, unlike the previous analysis of this dataset which used CSF as a reference region, and thus sensitivity correction was implicitly incorporated by the voxel-wise calibration, accounting for variable sensitivity of the radiofrequency receive head coil being used. Assumptions included a tissue T1 value of 1300 ms, an arterial blood T1 value of 1650 ms, and a blood-brain partition coefficient of 0.9. The labelling efficiency was set at 0.85 [1] . Slice timing effects were corrected by using a PLD of 1800 ms increasing for more superior slices with the slice delay of 30 ms for each. Partial volume effect correction (PVEc) was performed by BASIL [109] in OXASL using PV estimates supplied from FAST (volumetric pipeline) or Toblerone (surface-based pipeline) in the volumetric space. In the surface-based pipeline, (non-PVEc and PVEc) CBF maps were then projected onto the cortical surface. Finally, regional analysis was carried out for CBF maps in the native space in the volumetric pipeline. A threshold of 80% grey matter partial volume was used to define "pure" grey matter, where the mean GM CBF was calculated for the non-PVEc data. For PVEc data, the mean GM CBF was calculated and averaged across all voxels within the GM mask in which there was GM tissue. Examples of processed CBF maps in volumetric space were shown in Figure 3.2.3. Notably, zero-perfusion vertices were excluded and only non-zero values were calculated for the global and parcels.

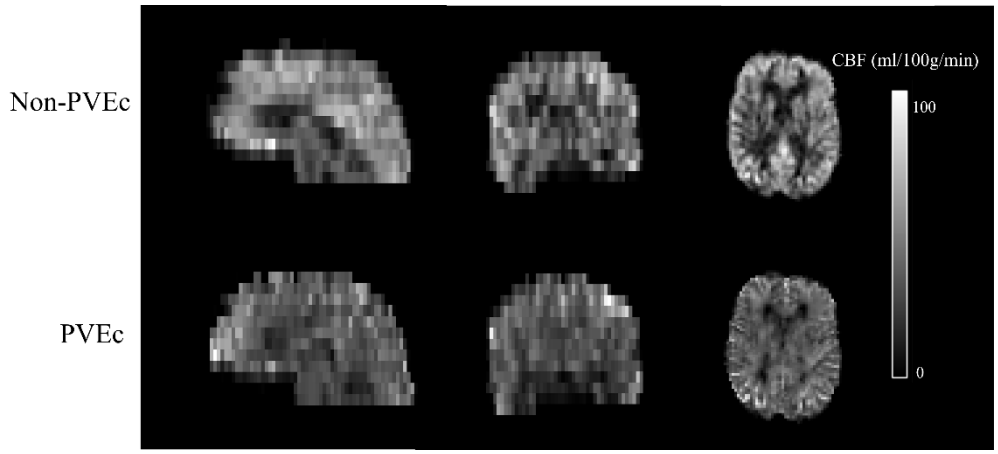


Figure 3.2.3 An example of processed CBF maps (in the native space) without PVEc (top) and with PVEc (bottom) from the volumetric pipeline. The planes are (from left to right):sagittal, coronal, and axial.

3.2.3.4 Mapping pCASL onto cortical Surface

To enable surface-based analyses, this study followed the processing procedure from the HCP fMRI Surface pipeline [155] . Firstly, non-PVEc and PVEc CBF maps in MNI volume space were mapped onto the midthickness surface in native mesh using the ribbon-constrained (RC) algorithm from the HCP workbench command (-volume-to-surface-mapping). Weighted partial volume maps were input to the method to better distinguish the contribution of voxels partially inside or outside the grey matter ribbon. Ultimately, the cortical CBF maps was resampled onto 32k_fs_LR mesh. The non-PVEc and PVEc ASL images from a participant projected

onto midthickness surfaces on 32k_fs_LR mesh in MNI volume space is shown in Figure 3.2.4.

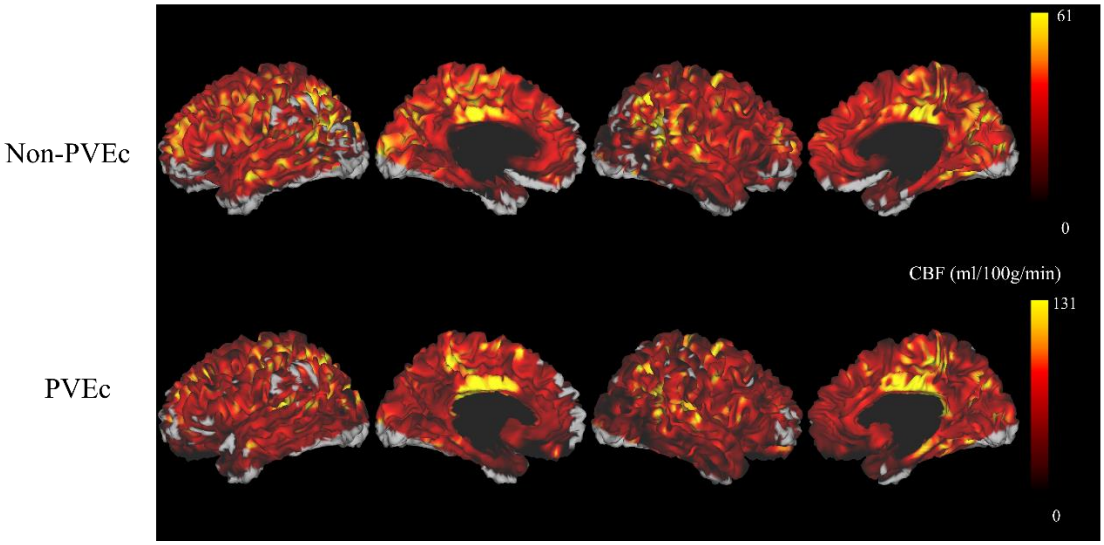


Figure 3.2.4 The same example of processed CBF maps (on to the 32k_fs_LR cortical midthickness surface) without PVEc (top) and with PVEc (bottom) from the surface-based pipeline.

3.2.4 Statistical analysis

Variations in CBF with age and sex were modelled using linear regression with ordinary least squares using python (version 3.11.8). Age was integrated into the models as a continuous variable, with sex serving as a covariate. Additionally, the models took into account potential interactions between age and sex. Statistical significance was set at $p < 0.05$. Notably, multiple comparisons were not corrected to remain the effects of ROIs as real as possible. In volumetric analysis, linear regressions were built for the whole brain grey matter and for ROIs derived from cerebral white matter, vascular territories [6] and the brain atlas including 18 regions (see Figure 3.2.5) used in UK Biobank imaging study neuroimaging analysis pipeline [171] . In surface-based analysis, models were calculated for the whole cortex and bilateral 33 cortical parcels from Desikan-Killiany atlas [172] . Normative reference values of non-PVEc and PVEc mean grey matter CBF are reported for both males and females at five-year intervals between the ages of 54 and 84, across the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles.

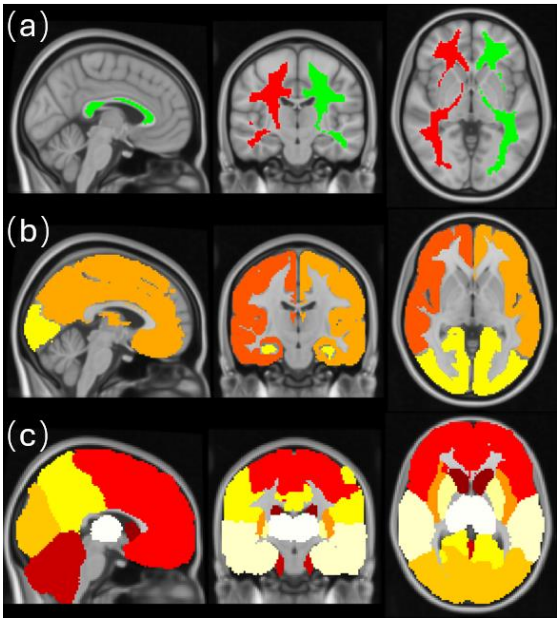


Figure 3.2.5 Atlases and ROIs used in volumetric analysis (in the standard space).

(a) left and right cerebral white matter; (b) vascular territories (Right Internal Carotid Artery Territory (RICA), Left Internal Carotid Artery Territory (LICA) and Vertebrobasilar Arteries territory (VBA); (c) 18 ROIs, used in UK Biobank : Caudate, Cerebellum, Frontal Lobe, Insula, Occipital Lobe, Parietal Lobe, Putamen, Temporal Lobe, Thalamus.

3.3 Results

423 subjects were successfully processed in this study. For each subject, we obtained mean and regional CBF values in volumetric regions and cortical parcels from non-PVEc and PVEc CBF maps.

3.3.1 Volumetric results

3.3.1.1 Effect of age on global CBF

Age-related GM CBF changes classified by sex and with/without PVEc are presented in Figure 3.3.1. For non-PVEc ASL data, mean GM CBF in the brain was 40.66 ± 8.5 ml/100 g/min; range: [14.98– 70.97 ml/100 g/min]. GM CBF decreased by 0.17 ml/100 g/min for each year of aging ($p<0.05$) and was on average 3.5 ml/100 g/min higher in females ($p<0.01$). In males, the decrease in GM CBF with age was significant ($p<0.05$) and was equivalent to a decrease of 17.1% across the age range tested from 41.21 ml/100 g/min in the youngest (54 years) to 36.08 ml/100g/min in

the oldest (84 years). However, the decline in GM CBF with age in females was not significant.

After PVEc, the mean global GM CBF in the brain was 48.56 ± 9.8 ml/100 g/min; range: [19.28– 82.6 ml/100 g/min]. GM CBF decreased by 0.18 ml/100 g/min for each year of aging ($p<0.05$) and was 3.8 ml/100 g/min higher in females ($p<0.05$). In males, the decrease in GM CBF with age was significant ($p<0.05$) and was equivalent to a decrease of 19.6% across the tested age range from 49.32 ml/100 g/min to 43.43 ml/100g/min. No significant variation was observed in females.

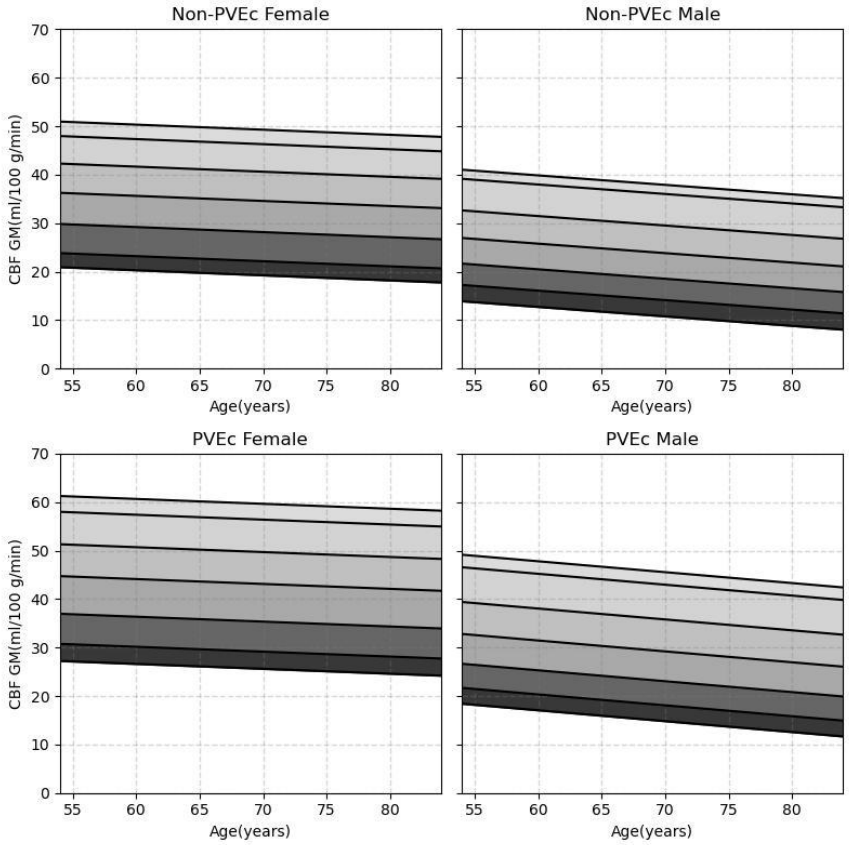


Figure 3.3.1 Age-related GM CBF changes by sex and with/without PVEc.

Top row non-PVEc results, and bottom row PVEc results. The age-related normative values for GM CBF are displayed across the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles from low to high. The decreasing trends were significant in males but not in females.

3.3.1.2 Effect of age on regional CBF

Age-related CBF variations in volumetric ROIs from non-PVEc and PVEc data were given in Table 3.3.1. Slopes of significance were marked with asterisks. In most

significant regions, CBF exhibited a decreasing trend, but an increase was only found in the left putamen in PVEc data.

Table 3.3.1 Age-related CBF variations in ROIs from the volumetric pipeline for non-PVEc and PVEc ASL.

	Non-PVEc			PVEc		
ROI	Slope	GM CBF	Sex Difference	Slope	GM CBF	Sex Difference
LICA	-0.12*	36.56	-4.25	-0.05	82.53	-4.4
RICA	-0.16**	40.28	-3.47	-0.17	83.97	-4.07
VBA	-0.09	40.18	-3.31	0.05	65.45	-3.71
Left Cerebral White Matter	-0.05	20.59	-2.5	-0.32**	63.8	-3.18
Right Cerebral White Matter	-0.07*	19.96	-2.38	-0.39**	62.57	-3
Left Caudate	-0.31**	24.17	-3.97	0.09	52.4	-3.23
Left Cerebellum	-0.06	31.35	-3.38	0.12	51.85	-2.92
Left Frontal Lobe	-0.07	37.04	-2.96	-0.11	86.02	-4.79
Left Insula	-0.18**	36.63	-2.32	-0.23*	65.1	-1.96
Left Occipital Lobe	-0.09	35.62	-3.88	0.13	72.69	-3.67
Left Parietal Lobe	-0.12*	36.15	-3.59	-0.08	89.02	-5.71
Left Putamen	0.09*	36.49	-1.57	0.24*	64.93	-0.5
Left Temporal Lobe	-0.08	32.85	-3.03	0.03	59.84	-3.32
Left Thalamus	-0.03	34.94	-4.27	0	82.29	-7.95
Right Caudate	-0.34**	23.55	-4.12	0.08	51.55	-3.45
Right Cerebellum	-0.06	31.18	-3.53	0.08	51.98	-3.23
Right Frontal Lobe	-0.12*	36.79	-2.95	-0.23	85.6	-4.6
Right Insula	-0.22**	35.89	-2.41	-0.30*	65.16	-2.46
Right Occipital Lobe	-0.1	36.56	-4.14	0.05	76.89	-4.75
Right Parietal Lobe	-0.13*	36.64	-3.38	-0.21	93.36	-5.44
Right Putamen	0.02	36.22	-1.57	0.1	65.66	0.33
Right Temporal Lobe	-0.1*	32.77	-2.61	-0.04	62.8	-2.45
Right Thalamus	-0.03	34.52	-4.29	0.06	81.52	-8.49

Slope Unit: ml/100g/min per year; CBF GM Unit: ml/100g/min; Sex difference(male-female) Unit: ml/100g/min; $P < 0.01$: **; $P < 0.05$: *.

3.3.2 Surface-based results

3.3.2.1 Effect of age on global cortical CBF

For non-PVEc CBF, the mean cortical GM CBF was 35.60 ± 7.34 ml/100 g/min; range: [19.1-47.5 ml/100 g/min]. GM CBF decreased by 0.22 ml/100 g/min for each year of aging ($p < 0.01$) and was on average 3.44 ml/100 g/min higher in females ($p<0.01$). For PVEc ASL data, mean cortical GM CBF was 67.09 ± 13.42 ml/100 g/min; range: [27.1-89.4 ml/100 g/min]. GM CBF decreased by 0.05 ml/100 g/min for each year of aging ($p>0.1$) and was on average 3.67 ml/100 g/min higher in females ($p<0.01$).

3.3.2.2 Effect of age on regional cortical CBF

Table 3.3.2 shows regional CBF GM variations with age of non-PVEc and PVEc data in cortical bilateral parcels from the Desikan-Killiany atlas before and after PVEc over all subjects. Before PVEc, parcels with significant GM CBF variations all exhibited decreases with age. After PVEc, the decreasing trend of GM CBF with age was no longer statistically significant in most parcels. Only seven parcels (caudal anterior cingulate, caudal middle frontal, pars opercularis, rostral anterior cingulate, superior frontal, superior parietal, and insula) still exhibited significant reduction with age, while significantly increasing CBF with age was found in the banks superior temporal, fusiform, and inferior temporal, regions.

Table 3.3.2 The cortical GM CBF variations with age with and without PVEc in Desikan-Killiany atlas over all subjects.

	Non-PVEc			PVEc		
ROI	Slope	GM CBF	Sex Difference	Slope	GM CBF	Sex Difference
Banks superior temporal sulcus	-0.17**	42.07	-4.14	0.32**	77.88	-1.08
Caudal anterior-cingulate cortex	-0.46**	29.93	-3.14	-0.3**	65.53	-2.74
Caudal middle frontal gyrus	-0.26**	44.38	-4.5	-0.21*	87.99	-5.56
Cuneus cortex	-0.24**	42.55	-4.95	0.1	74.26	-5.21
Entorhinal cortex	-0.06	20.11	-1.73	-0.04	36.7	-5.08
Fusiform gyrus	-0.06	27.18	-5.52	0.14*	53.33	-10.46
Inferior parietal cortex	-0.31**	43.1	-4.38	0.08	81.37	-1.23
Inferior temporal gyrus	-0.07	23.44	-4.66	0.12*	44.64	-5.47
Insula	-0.28**	36.01	-2.6	-0.29**	59.5	-2.75
Isthmus-cingulate cortex	-0.29**	37.8	-3.29	0	80.83	-7.53
Lateral occipital cortex	-0.25**	32.14	-6.65	-0.1	64.45	-7.77
Lateral orbital frontal cortex	-0.07*	27.14	-2.16	0.02	46.28	-2.71
Lingual gyrus	-0.12**	35.48	-5.05	0.08	68.74	-11.37
Medial orbital frontal cortex	-0.19**	29.08	-2.56	-0.08	44.18	-2.67
Middle temporal gyrus	-0.14**	37.19	-3.81	0.12	62.72	-2.92
Paracentral lobule	-0.16**	42.45	-3.24	0.07	81.99	-3.31
Parahippocampal gyrus	-0.02	28.06	-4.11	0.11	52.79	-8.05
Pars opercularis	-0.28**	41.68	-2.98	-0.16*	74.12	-2.95
Pars orbitalis	-0.06	23.41	-0.59	0.09	45.83	-0.91
Pars triangularis	-0.24**	38.4	-2.37	-0.07	66.17	-2.57
Pericalcarine cortex	-0.17**	41	-4.46	0.07	74.82	-4.82
Postcentral gyrus	-0.2**	42.45	-3.35	0	83.53	-3.66
Posterior-cingulate cortex	-0.49**	35.6	-2.25	-0.13	78.96	-2.68
Precentral gyrus	-0.17**	42.48	-2.75	0	85.72	-2.41
Precuneus cortex	-0.28**	42.62	-3.89	-0.01	77.61	-4.38
Rostral anterior cingulate cortex	-0.43**	39.39	-3.49	-0.21**	62.49	-4.09
Rostral middle frontal gyrus	-0.17**	39.58	-3.21	-0.03	67.2	-2.35
Superior frontal gyrus	-0.21**	39.44	-3.34	-0.18*	72.78	-4.18
Superior parietal cortex	-0.33**	38.98	-5.32	-0.22*	82.33	-5.4
Superior temporal gyrus	-0.2**	39.51	-2.56	0.07	69.61	-2.1
Supramarginal gyrus	-0.27**	41.45	-3.11	-0.01	77.33	-2.24
Temporal pole	-0.1*	19.23	-1.7	-0.02	31.21	-5.27
Transverse temporal cortex	-0.3**	48.44	-1.21	0.08	91.03	0.7

*Slope Unit: ml/100g/min per year; CBF GM Unit: ml/100g/min; Sex difference(male-female) Unit: ml/100g/min; P < 0.01: **; P < 0.05: *.*

The significant slopes of CBF changes with age from surface-based results, before and after PVEc, were mapped on corresponding cortical parcels (in Figure 3.3.2) to ease the visualization of identified trends and for further investigation on the spatial continuity.

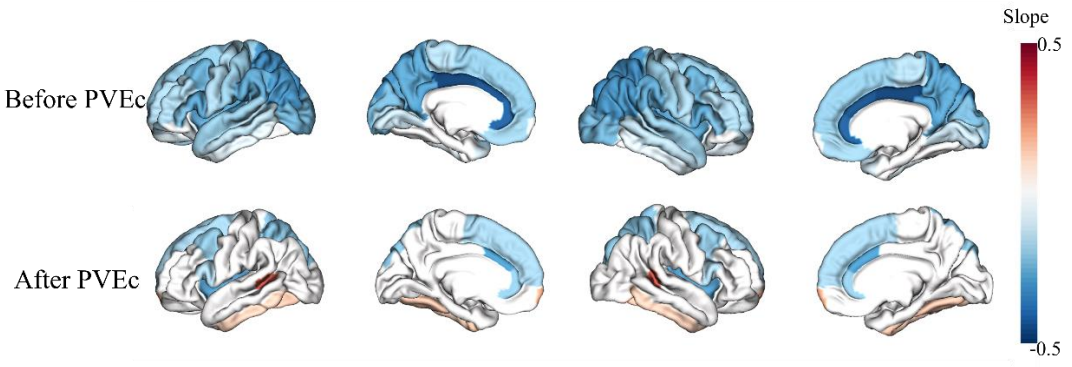


Figure 3.3.2 The slopes of age-related CBF changes from the surface-based results mapped on to corresponding parcels of the cortex. Slope unit: ml/100 g/min per year.

3.4 Discussion

This study investigated the regional changes in perfusion with age and explored the influence of changes in anatomy with age on these observations by attempting to correct for variation in grey matter volume. To our knowledge, this is the first study using both volumetric and surface-based analyses with the correction of partial volume effects with ASL MRI perfusion in a large cohort.

Perfusion images, especially ASL, are conventionally analysed using the volumetric method which is susceptible to partial volume effects. In this study, we mapped the ASL image onto the cortical surface to provide a better representation of cortical grey matter anatomy [162] 165]. Furthermore, we compared PVEc and non-PVEc results, following the suggestions from [4] , specifically to explore the role of PVEc and thus control for apparent changes in perfusion that are actually due to alterations in anatomy.

The former study on the TILDA ASL data [163] reported a mean global GM CBF (36.5 ± 8.2 ml/100 g/min) and decreased by 0.2 ml/100 g/min per year without PVEc. In contrast, this study found a slightly higher mean GM CBF in the whole brain (40.66 ± 8.5ml/100 g/min) and a lower CBF decreasing slope by 0.17 ml/100 g/min per year before PVEc. The reason for the observed differences could be largely

due to the different calibration methods employed, the use of voxel-wise calibration replacing the CSF reference region calibration in the former study, consistent with the recommendations of [1] . Pinto J et al. reported higher CBF values in voxel-wise calibration compared with CSF reference region using single-PLD pCASL data , which is consistent with our results [26] . Prior research has employed a range of perfusion imaging techniques to explore normative non-PVEc GM CBF values for different age groups. A study by Jefferson et al., following a similar protocol to this, reported mean whole brain CBF values of 37.3 ± 7.1 ml/100 g/min for 270 adults with an average age of 73 ± 7 years [173] . In a smaller study with different age groups using 3T pASL-MRI, Chen et al. reported higher mean cortical CBF values of 52.6 ± 9.3, 52 ± 10.7, and 42.7 ± 8.8 ml/100 g/min in young, middle-aged, and older groups respectively [150] . Biagi L, et al., using continuous ASL found a mean GM CBF 58.4 ml/ 100 g/minute for the 21 adults (mean age 40 ±15 years) [174] . After PVEc, mean GM CBF was found to be higher compared to non-PVEc GM CBF which was consistent with previous comparisons with and without PVEc [175] . There are few existing studies reported global perfusion with PVEc. For instance, a study by Meltzer CC et al. reported mean cortical CBF using PET for younger (62 ± 10 ml/100g/min) and older (62 ± 10 ml/100g/min) groups with PVEc, which was close to our results despite the use of a different perfusion technique [176] . Preibisch et al. using pASL MRI reported global GM CBF values 40.9 ± 5.5 ml/100g/min and 42.0 ± 8.6 ml/100g/min for 19 young and 25 older adults respectively with PVEc[45] .

We found global decrease of GM CBF with age before PVEc in both volumetric and surface-based analysis, and the slopes were not statistically different when evaluated by t-test. After PVEc, we found a greater decline of GM CBF with aging in volumetric analysis, although the slope was not statistically significantly different from that found before PVEc. Since post PVEc GM perfusion values were larger than pre-correction, the greater slope observed might simply be due to this, hence we also calculated the slope scaled by the mean GM CBF before and after PVEc, and we found the former was bigger than the latter (ratio between them 1.11). The relative reduction in the scaled slope values would be consistent with PVEc removing a component of apparent perfusion reduction with age that is associated with partial

volume effects and changes in brain structure. These results would tend to support the view that there is a genuine decrease in perfusion in GM with age [150] -[152] .

After PVEc, the surface-based analysis indicated a negligible change in CBF with aging: a small decreasing slope was observed, but this was not statistically significant. This potentially contradicts the volumetric results. However, it might be a result of the current acquisition and analysis employed in this case, potentially limiting the generalizability of these findings to studies using more typical imaging resolutions. For example, the pCASL acquisition has a relatively large inter-slice gap of 1 mm, which may lead to undersampling of tissues and thus exacerbate PVE, as inter-slice gap may reduce the accuracy of registration, where anatomical structures might be mismatched between the high-resolution structural images and the lower-resolution CBF images, resulting in misestimation of tissue types within voxels. Thus ASL data with higher resolution and partial volume correction optimised for surface analysis would ideally be adopted in future work to examine this [140] .

Regarding regional age-related GM CBF variations in our study, in the volumetric analysis, we found regional GM CBF reductions with age in most ROIs before PVEc, in some ROIs this was no longer significant after PVEc. In surface-based analyses, the regional cortical GM CBF values were found to decrease with age before PVEc. After PVEc, seven parcels remain a decreasing trend with three parcels showing an increase in perfusion. Previous studies have also investigated regional CBF changes with age with and without PVEc [150] ,[178] -[181] . For example, Parkes et al. observed that age-related changes in grey matter without PVE correction were predominantly localized in the frontal cortex using CASL MRI [181] . Martin et al. found that non-PVEc CBF values decreased with age in several regions, including the cingulate, parahippocampal, superior temporal, medial frontal, and posterior parietal cortices bilaterally, as well as in the left insular and left posterior prefrontal cortices [182] . Lee et al. identified both decreased and increased regional non-PVEc CBF values, with the most common regions for decreased perfusion being the precuneus, superior temporal, and orbitofrontal, and for increased perfusion, the caudate, posterior cingulate, anterior cingulate, and amygdala [46] . Parkes et al. detected age-related non-PVEc GM CBF reduction in the anterolateral prefrontal cortex and in areas along the lateral sulcus and the lateral ventricle, bilaterally [181] . Zhang et al. observed that CBF without PVEc using pCASL demonstrated decreases

with age in the frontal and parietal regions and the cerebellum, alongside increases in the temporal and occipital areas [148] . Preibisch C et al. using pASL MRI reported age-related CBF decreases without PVEc in frontal, parietal, insular, cingulate, parahippocampal and caudate. Few prior studies reported cortical GM CBF with PVEc. For instance, Preibisch C et al. reported decreases in GM CBF after PVEc in the parietal cortex, cuneus and caudate [45] . Furthermore, Preibisch C et al. also observed increases of CBF after PVEc, similar to ours, in the lateral and medial temporal lobe such as hippocampus, the calcarine gyrus and the thalamus.

Previous studies have also highlighted a dissociation between regional CBF and structural alterations specific to normal aging. Moreover, they suggest that other factors might influence age-related perfusion changes. For instance, Chen JJ et al. reported that regions experiencing CBF reduction are largely distinct from those most affected by GM atrophy, indicating that hemodynamic and anatomical changes may differentially contribute to age-related cognitive decline [150] . In the study by Parkes et al., they suggest that, without significant medical conditions, healthy aging might not affect resting cortical perfusion [181] . Other research indicates that the observed reduction in perfusion could be attributed to progressive neuronal loss, reduced neuronal activity, and a decline in the synaptic density of brain neurons [183] ,[184] .

We also investigated sex differences in age-related CBF changes and found females exhibited higher global GM CBF values compared to males before and after PVEc ($p<0.01$), which was consistent with some other studies [185] ,[186] . Furthermore, previous studies have investigated the rate of global CBF decline between males and females with respect to normal aging [186] ,[187] . However, we cannot reach a related conclusion, since the variations of global CBF with age in females was not statistically significant.

One limitation regarding to our pCASL MRI protocol is that it applied the same labelling duration and delay for all participants, effectively presuming negligible impact from possible spatial variations in arterial transit time (ATT)—the time it takes for the arterial blood bolus to travel from the labelling plane to the imaging voxels. However, ATT can vary regionally between individuals due to different age groups, health conditions, and populations [1] . A PLD of 1800 ms was considered

the optimal choice among our cohort with a mean age of less than 70 years to obtain adequate tissue signals and to minimize intravascular signals [188] . Consequently, this could lead to potential underestimation or overestimation of CBF measurements across the cohort of a wide age range (54 to 84) in this study. The 8 mm slice thickness of ASL acquisitions was substantially large for attempting surface mapping, and this may lead to differences in representations from both pipelines. Furthermore, small patches of zero perfusion of CBF map were spotted on the cortex (for example, Figure 3.2.3). The reason may be that the cortex of CBF map was not perfectly aligned with the template surface since registrations of CBF map (from native space to structural space and finally to the standard space) were conducted using volumetric methods (i.e. linear and non-linear registrations), or the deficiency of CBF map itself. In addition, extremely high perfusion of some vertices was found on the cortex which may be related to macrovascular contamination that cannot be corrected for in the single-PLD ASL data. It is worth noting that cortical perfusion could fluctuate slightly if resampled on a cortical mesh of different resolutions due to interpolation, and thus we made efforts to avoid redundant registration/smoothing as much as possible.

For ROI analyses, the influences of age and sex on CBF were investigated using linear regression in 35 volumetric ROIs and on 33 bilateral cortical parcels. The threshold for statistical significance level was $p < 0.05$ without the correction of multiple comparisons to avoid missing significant results, however, this will increase the risk of Type I errors and false discovery rate given the number of regions reported and different analysis approaches adopted in this study. Since this study was exploratory, we have not implemented multiple comparison corrections and have simply reported all results as a reference for other work. Future studies will be required with focused hypotheses to draw firm conclusions about changes in specific brain regions.

3.5 Conclusion

In this chapter, partial volume effects were corrected on a large elder pCASL data so far with the proposed surface-based pipeline as well as a volumetric pipeline. These processing pipelines were expected to serve as a template for subsequent ASL data

processing. Age-related CBF changes with and without PVEc were demonstrated using the surface-based and the volumetric representation in cortical parcels and volumetric ROIs, respectively.

Before PVE correction, a global decrease in GM CBF with age was observed. However, after applying PVEc, the perfusion variations became more regionally specific. While these findings suggest regional changes in CBF linked to aging, further studies with focused hypotheses are necessary to draw definitive conclusions about specific brain regions and to address other limitations identified in this study. Based on the results in this chapter, the changes in CBF (both non-PVEc and PVEc) in regions might be useful as metrics for assessing the quality of CBF maps.

In the next chapter, we will explore the use of CBF values in global and selected regions as quality metrics for CBF maps, verifying their effectiveness through machine learning models. Additionally, other ASL-specific image quality metrics will be investigated for their potential to evaluate the quality of CBF maps, aiming to standardize QC protocols for ASL using these IQMs.

4 Quality Control in ASL MRI

4.1 Introduction

Arterial Spin Labelling (ASL) MRI is a powerful non-invasive imaging technique that has gained prominence in the study of brain function and the detection of neural disorders [189] [193] . ASL uniquely enables the quantification of cerebral blood flow (CBF) without the need for exogenous contrast agents by magnetically labelling the blood water as it flows into the brain. This method offers several advantages, including safety for repeated measurements, the ability to capture dynamic changes in brain activity, and suitability for various populations. ASL has been widely applied in neuroscience research for understanding brain physiology, assessing cognitive functions, and diagnosing conditions such as stroke, Alzheimer's disease, and other neurodegenerative disorders [194] . Abnormal regional brain perfusion can be used as the biomarker of underlying neural pathologies which makes ASL a critical tool for early detection and monitoring of brain disorders.

However, like other MRI techniques, ASL MRI faces challenges such as motion sensitivity and susceptibility to artifacts, which can affect image quality and lead to inaccuracies in CBF measurements. To address these issues and facilitate the translation of ASL to clinical application, significant advancements have been made [194] , including improvements in image quality [195] , reduced acquisition times [196] , and enhanced reliability and reproducibility across different centers and scanners . Standardized acquisition methods [1] and image processing techniques have also been developed, alongside specialized software and tools like BASIL [109] , ExploreASL [85] , ASLPrep [86] , ASL-MRICloud [197] , and OXASL [198] . Most current ASL studies [6] [163] ,[242] now rely on these automated pre-processing pipelines to ensure efficiency and reproducibility. These pipelines typically involve steps including motion correction, distortion correction, registration, and quantification. However, standardized QC procedures were still absent across these tools. QC is the process identifying and excluding outliers in the data, helping to distinguish genuine perfusion changes from artifacts or low-quality

images introduced during acquisition and processing, thereby ensuring the reliability of the results before any analysis is conducted.

Furthermore, one crucial step in ASL pre-processing is registration [68] ,[69] , which involves aligning ASL images with anatomical reference images and, subsequently, with the standard template images. This alignment is essential for accurate measurement of grey matter perfusion, as it requires the grey matter partial volume estimate from anatomical images to be precisely aligned with the ASL data. Proper registration not only improves the accuracy of perfusion measurements but also enhances other pre-processing and analysis steps, such as motion correction, distortion correction, region-of-interest (ROI) mapping, and group analysis across subjects and sessions.

Registration in ASL imaging is often challenging due to the low resolution of ASL images, especially when dealing with variations in brain tissue characteristics, different fields of view, brain locations, orientations, and anatomical features. Previous works [199] -[203] have proposed different algorithms for registration to minimize potential errors to make more accurate registrations. However, the quality of registration is hard to assess, as there is no ground truth or gold standard of a perfect registration. In a study by [104] , the quality of registration was manually scored and categorized by experts, providing a subjective assessment of how well the registration had been performed. While this approach offers valuable insights, it is time-consuming and prone to subjectivity. A number of metrics have been used for evaluating registration quality [204] -[206] , however, they have their own shortcomings. For example, anatomical landmarks and contour-based metrics utilize the alignment of corresponding anatomical structures after registration, which depends on the annotation of anatomical structures or definition of accurate segmentation [206] ,[207] ; consistency-based metrics expect the concatenation of transformation matrices of registration in both directions to be the identity matrix. This method does not measure registration accuracy but consistency and cannot infer the error [208] . Some studies have employed deep learning to estimate the registration quality [209] . These methods require a large amount of data to train the models, however, obtaining such extensive datasets is challenging for most MRI studies. In response to these challenges, similarity-based metrics have become popular for evaluating registration quality. These metrics assess how closely the

registered image matches the reference image using quantitative measures such as mutual information, cross-correlation, or intensity differences. While useful, these metrics can be sensitive to noise and low-intensity contrast, common challenges in ASL imaging [210] .

The consensus paper [1] outlined a manual approach for visual QC of ASL images, focusing on specific issues such as low labelling efficiency, global grey matter CBF values, and various artifacts. While this method addresses essential quality checks, it is labor-intensive, subjective, and highly dependent on the expertise of the raters, leading to potential inconsistencies and inefficiencies. This approach becomes particularly challenging with large-scale datasets like those from the UK Biobank or the Human Connectome Project [171] ,[231] , which include thousands of scans, making manual visual examination impractical. A later study by [104] attempted to standardize the QC process by introducing a visual scoring system for ASL data. This system evaluated CBF maps along with ancillary images like aBV and ATT maps, using a dual approach: contrast-based QC for anatomical structure contrast and artifact-based QC for common visual errors. Each aspect was scored on a scale from 0 to 2, with higher scores indicating better image quality. Although this method provided examples for different severity levels to guide raters, it still relied on manual scoring. While more structured than the approach outlined in the consensus paper, this method also faces challenges in handling large datasets.

Given the challenges associated with ASL data, automated QC tools are increasingly desired to reduce the burden of manual QC, minimize subjectivity, and enhance consistency across large datasets. In other MRI modalities, automated QC tools have been successfully integrated into workflows, such as those used in the UK Biobank [171] and MRIQC [83] for T1-weighted images, FSL EDDY [190] for dMRI, and FMRIPrep [211] and AFNI [84] for fMRI images. These tools effectively address common MRI quality issues like motion artifacts, signal dropouts, and geometric distortions. However, they are typically not specifically designed to handle the unique challenges and requirements for ASL MRI. Recently, ASL tools like ExploreASL [85] and ASLPrep [86] have been developed for ASL image processing and analysis, with integrated QC components. These tools generate QC reports that assist in the visual QC process, providing a more streamlined and standardized approach. However, their QC methods are dependent on the intermediate and final

outputs of their workflows, limiting their applicability to ASL data processed by other tools. OXASL [198] is a package designed for performing Bayesian analysis of ASL MRI data, serving as a mostly Python-based replacement for the BASIL FSL. Its efficiency, reliability, and ease of use have been validated by previous studies [212] [213] , while it lacks an automated QC function.

Machine learning offers a promising alternative for ASL quality control regardless of both manual visual scoring and the limitations of individual QC metrics. Machine learning can incorporate multiple computed QC metrics, capturing the complex relationships between them, to make more informed decisions about the quality of data. Instead of relying on a single metric or subjective visual assessment, machine learning models can analyse a combination of metrics to identify outliers and assess the overall quality of the images. This approach not only reduces the subjectivity and labour involved in manual QC but also enhances consistency and accuracy across large datasets. This approach has already shown promise in other medical imaging modalities. For instance, machine learning models have been used to detect outliers and assess image quality in large datasets, such as those in the UK Biobank [171] . Although these applications have primarily focused on other types of MRI data, the principles can be adapted to ASL imaging, where machine learning could offer a more scalable and objective alternative to both visual scoring systems and traditional QC metrics. In studies using other modalities by [103] ,[171] ,[214] , machine learning has been successfully employed to predict and exclude outliers by integrating various quality metrics. Applying a similar approach to ASL data could help recognize patterns indicative of potential issues, thereby enabling the automated exclusion of problematic images. This would make the process more efficient, reliable, and scalable, especially in large-scale studies.

In this chapter, we explored image quality metrics specifically for ASL MRI to facilitate QC on large datasets. These ASL QC metrics included signal quality metrics, perfusion quality metrics, and registration quality metrics. Additionally, interactive QC reports were generated following the similar structure of [190] : the group-level report presents the distribution of ASL QC metrics across the dataset, enabling the identification of abnormal values (potential outliers) through user-defined thresholds, while the individual-level report details pre-processing results from OXASL to assist in troubleshooting suspicious images. Moreover, machine

learning was employed to assess the effectiveness of each proposed QC metric, ultimately leading to the development of an automated QC tool for ASL data using these ASL QC metrics.

4.2 Theory

4.2.1 Signal Quality Metrics

MR images are inherently contaminated by background noise, arising from stochastic fluctuations in the imaging systems (such as thermal motion of electrons in the receiver coil) or the tissues being imaged (such as physiological fluctuations for fMRI) [216] ,217]. The MRI signal is commonly detected in quadrature using receiver coils, which records real and imaginary components. Each component is assumed to have zero-mean Gaussian noise, with independent white noise contamination. When computing the magnitude image, the noise follows a Rician distribution, rather than a Gaussian one, as it arises from the square root of the sum of squared real and imaginary components as complex numbers [217] . This results in a noise bias, particularly in low signal regions, where the noise bias becomes more prominent. This bias, often referred to as the noise floor, results in increased variance and systematic overestimation of low-intensity signals. Consequently, Rician noise can affect MRI data quality by reducing contrast, distorting quantitative measurements, and introducing bias in parameter estimates [218] . In ASL, where perfusion signals are inherently low, this noise-induced bias can significantly impact the estimation of signals from blood water and thus bias the CBF quantification. Moreover, ASL signals are highly susceptible to various factors such as blood flow, T1 relaxation times of blood and tissue, and the transit time of blood from the labelling plane to the imaging region [219] . Given these challenges, the measurement of ASL signals is a critical preliminary step in the quality control process, aiming to assess and ensure the reliability of the perfusion data. This section focuses on investigating signal quality metrics, providing a foundation for subsequent quality control measures.

4.2.1.1 Signal-to-noise Ratio

SNR is a fundamental measure used to compare the level of the desired signal to the level of background noise, serving as the benchmark for evaluating the overall efficacy of the imaging process [220] . SNR is calculated on the perfusion-weighted images (that is the average of differences between label-control image pairs) by comparing the signal intensity from a region of interest, typically grey matter (GM), to the noise level, which can be referenced against white matter (WM), cerebrospinal fluid (CSF), or even air. A higher SNR generally indicates a clearer and more reliable image. However, a high SNR does not always correspond to the quality of the perfusion signal in ASL images. For example, an image with strong overall signal intensity (high SNR) may still have poor perfusion contrast due to issues like inadequate labelling or delayed blood arrival. SNR can be calculated using the following formula:

$$SNR = \frac{S}{\sigma_{noise}} = \frac{Mean\ Intensity_{GM}}{SD_{background}} \quad \textit{Equation 4.2.1}$$

where $Mean\ Intensity_{GM}$ is defined as the mean intensity of grey matter, $SD_{Background}$ uses the standard deviation of the intensities of other type of tissues (WM or CSF) or air.

4.2.1.2 Temporal Signal-to-noise Ratio

Temporal SNR (tSNR) extends the concept of SNR by incorporating the temporal dimension and evaluates the stability of the signal over time. In ASL, tSNR is calculated after motion correction using the ASL difference images, which are derived from the sequence of subtracted label-control image pairs. A high tSNR indicates that the signal remains consistent across multiple acquisitions, which is important to ASL that requires repeated acquisition of label-control image pairs. The tSNR can be influenced by factors such as motion artifacts and physiological noise.

The tSNR is typically formulated as:

$$tSNR = \frac{\overline{S_t}}{\sigma_t} = \frac{Mean_{GM\ over\ time}}{SD_{GM\ over\ time}} \quad \textit{Equation 4.2.2}$$

Where $Mean_{GM\ over\ time}$ and $SD_{GM\ over\ time}$ are the mean and the standard deviation of the ASL difference images across time points within in the GM mask, respectively.

4.2.1.3 Contrast-to-noise Ratio

Contrast-to-Noise Ratio (CNR) measures the difference in signal intensity between two tissue types, such as grey matter and white matter [221] , relative to the noise level. In ASL MRI, CNR is calculated using the ASL difference images, after motion correction, to obtain the contrast of perfusion signal by comparing the signals within GM masks and WM masks. A higher CNR suggests better differentiation between the two types of tissues, allowing better visualization of anatomical details. CNR is defined as:

$$CNR = \frac{\Delta S}{\sigma_{noise}} = \frac{abs(S_{signal} - S_{background})}{\sigma_{noise}} \qquad \textit{Equation 4.2.3}$$

In ASL, S_{signal} is the mean signal intensity in grey matter, $S_{background}$ is the mean signal intensity in background which can be either CSF or WM; σ_{noise} is the standard deviation of signals in the background; abs is to calculate the absolute value.

4.2.1.4 Temporal Contrast-to-noise Ratio

Temporal Contrast-to-Noise Ratio (tCNR), similar to tSNR, measures the consistency of contrast over time, relative to the temporal noise, thereby quantifying dynamic changes in perfusion signals. This metric is particularly relevant in ASL MRI, where the perfusion signals are acquired across different time points. Temporal CNR helps in evaluating the reliability of perfusion contrast throughout the imaging session, ensuring that the detected differences in perfusion are not due to noise or temporal instability. However, tCNR is susceptible to motion artifacts or physiological noise. Temporal CNR can be calculated using the following formula:

$$tCNR = \frac{\overline{\Delta S_t}}{\sigma_t} = \frac{Mean_{CNR\ over\ time}}{SD_{CNR\ over\ time}} \qquad \textit{Equation 4.2.4}$$

Where $\overline{\Delta S_t}$ is the mean CNR of the ASL difference map; σ_t is the standard deviation of CNR values from the ASL difference map.

4.2.2 Perfusion Quality Metrics

In the brain, GM is more sensitive to perfusion due to its higher metabolic activity, making it the primary tissue of interest for ASL. In this section, CBF, the quantitative perfusion measure, and its spatial distributions in the brain and specific ROIs were investigated as key indicators of perfusion quality in ASL data.

4.2.2.1 Gray Matter Cerebral Blood Flow

Gray matter contains the majority of neuronal cell bodies, axon terminals, and dendrites, all of which have a higher metabolic rate compared to white matter. Consequently, it is of primary interest for ASL MRI because this higher metabolic activity makes grey matter more sensitive to detecting changes in brain perfusion [222] . Furthermore, grey matter is involved in various critical brain functions, including sensory perception, memory, and decision-making. Accurate measurement of GM CBF can provide valuable insights into these cognitive and functional processes.

The consensus paper [1] reported a general GM CBF range of 40-100 ml/min/100ml. Some studies [163] ,[186] reported a narrower range, suggesting a typical perfusion of 60 ml/100g/min in GM and 20 ml/100g/min in WM. Furthermore, the normative CBF values changing with age and sex were reported in studies [150] ,[163] ,[223] , which provide a better guide to detect outliers with abnormal GM CBF. These reference values are invaluable for detecting abnormalities and potential diseases by comparing individual GM CBF metrics against established baselines. Such comparisons are crucial for identifying deviations that could indicate pathological conditions. Therefore, GM CBF is not only a key indicator of brain health but also a critical metric for assessing the quality of ASL data.

4.2.2.2 Spatial Coefficient of Variation

The arterial transit time (ATT) is a common error source leading to artifacts in ASL imaging [51] . In clinical settings, diseases such as stroke, tumors, and vascular abnormalities can affect ATT. A short delay in this transit time hinders the complete delivery of labelled blood water, while a long delay leads to signal decay and thus

impairs SNR. Additionally, ATT varies between participants across different brain regions, and among various pathological tissues. Hence, ATT provides valuable insights into the vascular health and hemodynamics of the brain.

Previous studies have found that the spatial distribution of ASL signals can be useful in inferring ATT differences across single-PLD CBF maps [225] -[228] . Specifically, if all ASL tracer is delivered to the tissue after the post-labelling delay (PLD), the ASL signal intensities will be spatially homogeneous within the tissue. Conversely, if the tracer is not fully delivered, the ASL signal intensities will vary markedly between the vascular and tissue regions (see Figure 4.2.1 from [228]). Therefore, the spatial coefficient of variation (SpCoV) can be used as a proxy for ATT to assess the efficiency of ASL acquisition [228] . High SpCoV values suggest regions where the perfusion signal is unstable, potentially due to artifacts or other acquisition issues, thereby providing an indirect measure of ATT that is crucial for optimizing ASL MRI studies. SpCoV for a ROI was defined as:

$$SpCoV = \frac{\sigma_{CBF}}{\mu_{CBF}} \times 100\% \qquad \textit{Equation 4.2.5}$$

where σ_{CBF} is the standard deviation of CBF within the ROI and μ_{CBF} is the mean CBF within the same ROI.

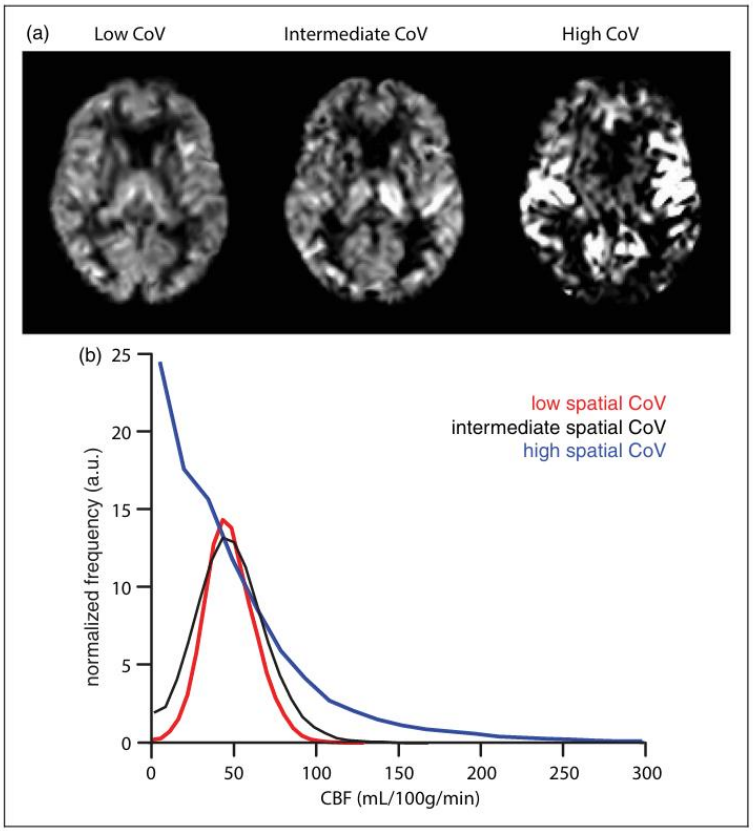


Figure 4.2.1 CBF maps of different (low, intermediate, and high) spatial coefficient of variation from [228] . (a): CBF maps; (b): histograms of voxel normalized frequency from CBF maps.

4.2.2.3 Perfusion in Arterial Vascular territories and cerebral white matter

Inefficient labelling is one of the most common artifacts in ASL, often resulting from factors such as vessel tortuosity or susceptibility variations (as discussed in Chapter 2). This artifact typically manifests as decreased signals in specific arterial vascular territories of CBF maps, making it essential to monitor perfusion within these territories to identify and address potential issues.

An arterial vascular territory atlas, as detailed in [6] and illustrated in Figure 3.2.5 (b), includes the territories supplied by the left and right internal carotid arteries as well as the vertebrobasilar arteries. In normal ASL images, the CBF values across these vascular territories should be relatively similar, and significant deviations could indicate labelling inefficiencies or other artifacts.

Furthermore, recent studies have shown that ASL can reliably measure white matter [171] perfusion when appropriate imaging settings are employed [229] , despite the lower SNR in white matter due to its lower blood flow and prolonged ATT compared to grey matter. The cerebral white matter regions were derived from the work (shown in Figure 3.2.5 (a)). Additionally, the ratio of CBF between grey matter and white matter has been reported to range from 1.6 to 4.6 [229] , may serve as a useful parameter for QC in ASL studies, as significant deviations from this range could indicate potential issues with image quality or data accuracy. Moreover, significant discrepancies in CBF between the left cerebral white matter (LCWM) and right cerebral white matter (RCWM) regions, CBF values falling outside the expected range, or an extreme perfusion ratio between GM and WM, can indicate potential issues with the ASL images.

4.2.3 Registration Quality Metrics

Assuming that registration is correctly performed for the majority of ASL CBF maps, which were aligned to the same reference image (i.e., the MNI template [70]) through intermediate structural spaces, the degree of similarity between a registered ASL image and the reference image can serve as an indicator of registration quality.

Extremely low similarity values should be the primary focus, as they may indicate potential registration failures. In this section, metrics measuring the similarity between images are introduced to evaluate ASL registration quality based on the aforementioned assumption.

Least Squares

The least squares (LS) cost function is one of the simplest and most commonly used methods for image registration. It measures the similarity between two images by calculating the sum of the squared differences (SSD) between corresponding pixel or voxel intensities. Mathematically, it is expressed as:

$$LS = \sum_{i=1}^{n}(I_{ASL}(i) - I_{Template}(i))^2 \qquad \textit{Equation 4.2.6}$$

where $I_{ASL}(i)$ and $I_{Template}(i)$ represent the intensities of the ASL and template images, respectively, at the i-th voxel. The lower the LS value, the better the alignment between the two images. However, this method is sensitive to intensity variations, noise, and outliers, making it less robust for ASL MRI registration, especially in cases where image intensity differences do not strictly correspond to misregistration.

Normalized Correlation

Normalized Correlation (NC) is another metric used to assess the similarity between the registered ASL image and the template. Unlike LS, NC is less sensitive to global intensity differences between images. The NC between two images is calculated as:

$$NC = \frac{\sum_{i=1}^{n}(I_{ASL}(i)-\bar{I}_{ASL})(I_{Template}(i)-\bar{I}_{Template})}{\sqrt{\sum_{i=1}^{n}(I_{ASL}(i)-\bar{I}_{ASL})^2 \sum_{i=1}^{n}(I_{Template}(i)-\bar{I}_{Template})^2}} \qquad \textit{Equation 4.2.7}$$

where $\bar{I}_{ASL}$ and $\bar{I}_{Template}$ are the mean intensities of the ASL and template images, respectively. NC values range from -1 to 1, with 1 indicating perfect correlation (i.e., perfect alignment). NC is more robust than LS in handling global intensity shifts, but it may still struggle with significant intensity non-uniformities.

Correlation Ratio

The Correlation Ratio (CR) is a more sophisticated metric that measures the functional relationship between the intensities of the registered ASL image and the template image. It accounts for non-linear intensity relationships, making it more robust in scenarios where the intensity distributions between the images differ. CR is defined as the ratio of the variance of the mean intensity values to the total variance of the intensity values:

$$CR = \frac{\sum_{i=1}^{n}(I_{ASL}(i)-\bar{I})(I_{Template}(i)-\bar{I})}{\sum_{i=1}^{n}(I_{ASL}(i)-\bar{I})^{2}(I_{Template}(i)-\bar{I})^{2}} \qquad \textit{Equation 4.2.8}$$

where $I_{ASL}(i)$ and $I_{Template}(i)$ represent the intensities of the ASL and template images, respectively, at the i-th voxel; $\bar{I}$ is the mean of ASL and template images. CR is effective in capturing complex intensity relationships, making it a valuable tool for evaluating registration quality when the intensity distributions between the ASL and template images are not linearly related.

Mutual Information

Mutual Information (MI) is frequently used in medical imaging for multi-modal image registration, where it helps to align images from different imaging modalities by maximizing the shared information between them. ASL data is often compared with structural MRI scans (such as T1-weighted images) for registration purposes. Given the different contrasts between ASL images and structural MRI, traditional methods like sum of squared differences might not be effective. Mutual Information, however, can align these multi-modal images by focusing on how information is shared between them, regardless of their differing contrast properties.

The MI can be formulated as:

$$MI = \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} P(I_{ASL}(i), I_{Template}(j)) log \frac{P(I_{ASL}(i), I_{Template}(j))}{P(I_{ASL}(i))P(I_{Template}(j))} \qquad \textit{Equation 4.2.9}$$

where $P(I_{ASL}(i), I_{Template}(j))$ is the joint probability distribution of the pixel intensities in the ASL and template images, and $P(I_{ASL}(i))$ and $P(I_{Template}(j))$ are their marginal distributions. Higher MI values indicate greater similarity, suggesting better registration. MI is effective in handling different imaging modalities but may be computationally intensive.

Normalized Mutual Information

Normalized Mutual Information (NMI) is an extension of MI that addresses some of its limitations by normalizing the mutual information score, making it less sensitive to changes in image overlap:

$$NMI = \frac{H(I_{ASL})+H(I_{Template})}{H(I_{ASL},I_{Template})} \qquad \textit{Equation 4.2.10}$$

where $H(I_{ASL})$ and $H(I_{Template})$ are the entropies of the ASL and template images, and $H(I_{ASL}, I_{Template})$ is their joint entropy. NMI ranges from 0 to 2, with higher values indicating better alignment. NMI is particularly useful for comparing images of different sizes or in situations where the overlap between images is variable.

4.3 Experiments

4.3.1 Dataset

This study used the same ASL dataset as described in Chapter 3, namely The Irish Longitudinal Study on Ageing (TILDA) [166] . Chapter 3 focused on investigating normative CBF decline in an elderly population, excluding outliers and artifacts. In contrast, the current study aims to identify outliers, retaining all available subjects for analysis. The non-PVEc CBF maps in the native space, processed with the volumetric ASL pipeline in Chapter 3, were employed for quality control in this chapter, encompassing a total of 474 subjects.

The CBF maps in the native space were reviewed for quality control purposes by two individuals from the prior study [163] , who identified and labelled artifacts. A total of 30 subjects were marked as outliers based on the types of artifacts present (see Figure 4.3.1 for examples). The artifact types included poor labelling efficiency (27 subjects), low contrast (1 subject), motion (7 subjects), poor signal (1 subject), and

delayed arrival (6 subjects), with some outliers exhibiting multiple artifact types.

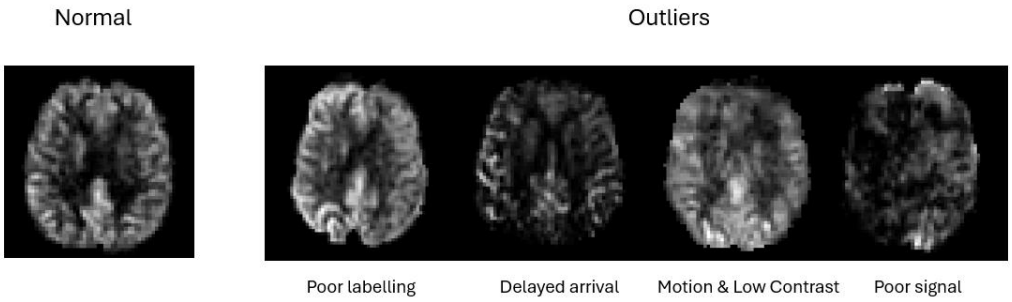


Figure 4.3.1 Normal (left) and outliers (right 4) examples in the ASL native space from the dataset.

4.3.2 Methods

Image quality metrics

Image quality metrics were calculated for each subject in the dataset, with details in Table 4.3.1. For the calculation of signal quality metrics, SNR and CNR were calculated based on the perfusion-weighted images (averaged ASL difference images), and tSNR and tCNR were computed based on the ASL difference (control-label) images. The noise region used eroded CSF masks from segmentation by FAST. To be detailed, SNR was calculated by comparing the GM signals (extracted from within the GM mask) of the perfusion-weighted images against these noise (CSF) signals. CNR was determined by comparing the signal between GM and WM to noise (CSF) signal within the perfusion-weighted images. Temporal metrics, including tSNR and tCNR, were computed using the SNR and CNR values from each ASL difference image, along with their standard deviations across the ASL difference images.

The perfusion quality metrics, including CBF values, along with the SpCoV, were calculated based on the CBF maps, with regions of grey matter, white matter, cerebral white matter, and arterial vascular territories for both non-PVEc and PVEc ASL data. Additional metrics included the CBF ratios between GM and WM, the comparison of CBF between the left and right carotid arteries, and the CBF ratio between the left and right cerebral white matter. CBF values falling outside the reasonable range, or significant discrepancies of CBF ratios can indicate potential issues with the ASL images.

The quality of image registration was evaluated by similarity metrics between the registered CBF images (aligned to the structural space and then to the standard space) and the MNI152 standard template [70] . These metrics included Least Squares (LS), Normalized Correlation (NC), Mutual Information (MI), Normalized Mutual Information (NMI), and Correlation Ratio (CR). Notably, these metrics were computed as a cost function of registration using FLIRT FSL post-registration, and thus this would not influence the effectiveness of registration.

Table 4.3.1 The proposed features used to evaluate ASL data quality.

Feature Name	Description
Signal quality metrics	
SNR	(Mean signal intensity in GM)/(Mean signal intensity in CSF) from perfusion-weighted image
tSNR	(Mean signal intensities in GM)/(temporal standard deviation) from the ASL difference (control-label) images
CNR	(Mean signal intenty in GM - Mean signal intensity in WM)/(Mean signal intensity in CSF) from perfusion-weighted image
tCNR	(Mean CNR)/(temporal standard deviation of CNR) from the ASL difference (control-label) images
Perfusion Quality Metrics	CBF=Cerebral blood Flow; PVEc=Partial Volume Effects correction;SpCov=Spatial Coefficient of variation=(mean CBF/standard deviation of CBF) within a ROI.
NonPVC GM CBF	CBF in gray matter from the non-PVEc CBF map
NonPVC WM CBF	CBF in white matter from the non-PVEc CBF map
NonPVC RCWM CBF	CBF in right cerebral white matter from the non-PVEc CBF map
NonPVC LCWM CBF	CBF in Left cerebral white matter from the non-PVEc CBF map
NonPVC VBA CBF	CBF in vertebrobasilar artery from the non-PVEc CBF map
NonPVC RICA CBF	CBF in right internal carotid artery from the non-PVEc CBF map
NonPVC LICA CBF	CBF in left internal carotid artery from the non-PVEc CBF map
NonPVC GM SpCov	SpCov in gray matter from the non-PVEc CBF map
NonPVC WM SpCov	SpCov in white matter from the non-PVEc CBF map
NonPVC RCWM SpCov	SpCov in right cerebral white matter from the non-PVEc CBF map
NonPVC LCWM SpCov	SpCov in Left cerebral white matter from the non-PVEc CBF map
NonPVC VBA SpCov	SpCov in vertebrobasilar artery from the non-PVEc CBF map
NonPVC RICA SpCov	SpCov in right internal carotid artery from the non-PVEc CBF map
NonPVC LICA SpCov	SpCov in left internal carotid artery from the non-PVEc CBF map
PVC GM CBF	CBF in gray matter from the PVEc CBF map
PVC WM CBF	CBF in white matter from the PVEc CBF map
PVC RCWM CBF	CBF in right cerebral white matter from the PVEc CBF map
PVC LCWM CBF	CBF in Left cerebral white matter from the PVEc CBF map
PVC VBA CBF	CBF in vertebrobasilar artery from the PVEc CBF map
PVC RICA CBF	CBF in right internal carotid artery from the PVEc CBF map
PVC LICA CBF	CBF in left internal carotid artery from the PVEc CBF map
PVC GM SpCov	SpCov in gray matter from the PVEc CBF map

PVC WM SpCov	SpCov in white matter from the PVEc CBF map
PVC RCWM SpCov	SpCov in right cerebral white matter from the PVEc CBF map
PVC LCWM SpCov	SpCov in Left cerebral white matter from the PVEc CBF map
PVC VBA SpCov	SpCov in vertebrobasilar artery from the PVEc CBF map
PVC RICA SpCov	SpCov in right internal carotid artery from the PVEc CBF map
PVC LICA SpCov	SpCov in left internal carotid artery from the PVEc CBF map
Registration quality metrics	
Least Square	Least square (or Sum of the Squared Differences) between the CBF map registered to the MNI space and the MNI template image
Norm Correlation	Normalized correlation between the CBF map registered to the MNI space and the MNI template image
Mutual Information	Mutual information between the CBF map registered to the MNI space and the MNI template image
Norm Mutual Information	Normalized mutual information between the CBF map registered to the MNI space and the MNI template image
Correlation Ratio	Correlation ratio between the CBF map registered to the MNI space and the MNI template image
Symmetric metrics	
NonPVC LICA/RICA	LICA CBF to RICA CBF ratio from the non-PVEc CBF map
NonPVC LCWM/RCWM	LCWM CBF to RCWM CBF ratio from the non-PVEc CBF map
NonPVC GM/WM	GM CBF to WM CBF ratio from the non-PVEc CBF map
PVC LICA/RICA	LICA CBF to RICA CBF ratio from the PVEc CBF map
PVC LCWM/RCWM	LCWM CBF to RCWM CBF ratio from the PVEc CBF map
PVC GM/WM	GM CBF to WM CBF ratio from the PVEc CBF map

The normalization used in this chapter is done by min-max feature scaling to restrict the range of values and to avoid arbitrary values. Furthermore, A CSV file was generated corresponding to these image quality metrics for each subject in the dataset, which can be used for subsequent QC research.

Quality Control Reports

QC reports were developed to facilitate the screening of ASL images and to detect common failure examples in the dataset, conducted at both the group and individual levels. These reports were generated using the processed images from OXASL and implemented using python with Jupyter Notebook. Furthermore, interactive controls were provided with the QC reports allowing the flexible screening process to their specific requirements.

Group-Level QC

The group-level QC report provided an overview of the entire dataset by displaying scatter plots alongside box plots for all ASL QC metrics. Additionally, interactive widgets were developed with the plots, allowing to filter subjects by user-specified metric and user-defined thresholds. This interactive functionality facilitates the identification of potential outliers across the group, allowing users to quickly pinpoint images that deviate from the norm and may require further examination, which can then be conducted using the individual-level QC report.

Individual-Level QC

To streamline the screening process for individual images, individual-level QC reports were generated. These reports presented key images and relevant supporting information by OXASL, making it easier to assess the quality of key steps of processing. The subject to be presented can be selected using a dropdown menu, allowing users to easily navigate between different subjects in the dataset. The individual-level report was divided into three sections based on the spatial context of the data:

1. **Structural Space QC**: The first section focused on evaluating the brain extraction and tissue segmentation of the T1w image in the structural space. Interactive plots allow users to view the brain-extracted T1w image along with GM/WM masks or contours. The images are presented in axial, sagittal, and coronal planes, with adjustable slices, enabling users to thoroughly inspect the segmentation quality.
2. **ASL Native Space QC**: This section examined the ASL images in the native ASL space. It presents perfusion-weighted images before and after motion and distortion correction. Additionally, CBF maps were displayed, where users could utilize dropdown menus to control the display of non-PVEc or PVEc GM CBF maps. These maps could be further customized by selecting different calibration methods, such as voxelwise calibration, or using WM or CSF as reference regions. Users also had the option to overlay the GM or WM mask onto the CBF maps, allowing for a more detailed and interactive examination of the data. This flexibility aids in comparing the effects of

various corrections and calibration methods on the quality of ASL data. Regional analysis was provided to present CBF and SpCoV in ROIs. Motion estimates were also presented using FSL MCFLIRT [71] which help in evaluating the extent of subject movement of the raw ASL data with the middle volume as an initial template image showing the rotation and translation series in three dimensions.

3. **Standard Space QC**: The final section was designed to examine the alignment of ASL data to the standard space, which was crucial for ROI mapping in group analysis. To this end, the template image, T1w image, and ASL image were all displayed within the standard space.

Quality Control Using Machine Learning

To evaluate the effectiveness of the proposed QC metrics and develop an automated QC tool for ASL data, machine learning was applied using the labelled data.

Support Vector Machines (SVMs) [121] were widely utilized for classification tasks, functioning by identifying the optimal hyperplane that separates two classes with the goal of maximizing the margin between the nearest data points from each class. While many potential hyperplanes could serve this purpose, SVMs seek the one that provides the largest margin, thereby ensuring the most robust decision boundary. Additionally, the kernel trick allows SVMs to handle non-linearly separable data by implicitly mapping the input features into a higher-dimensional space where a linear hyperplane can be identified.

In this study, the ASL dataset was notably unbalanced, a common challenge in medical imaging, with only 30 subjects (6.3%) out of 474 being labelled as outliers. Unbalanced datasets can cause most machine learning models to bias towards the majority class. However, SVMs focus on the boundaries between classes rather than the overall data distribution, making them particularly effective for handling unbalanced data. Moreover, for ASL quality metrics, extreme high or low values typically indicate errors and often lie at the two sides of data distribution, making it difficult to filter them using a simple linear decision boundary. The kernel trick of SVMs is well-suited to address this issue, as it can effectively separate outliers by

finding a more complex decision boundary in a higher-dimensional space. Therefore, SVMs are theoretically an excellent choice for this application.

The dataset consisted of 474 subjects, comprising 444 normal images and 30 outliers. It was randomly divided into two parts: 70% (306 normal images and 25 outliers) were used for model training, while the remaining 30% (138 normal images and 5 outliers) were reserved for testing.

All features were computed for the ASL images in the native space. The features used for classification included signal-related metrics (4 features), CBF, and SpCoV in GM, WM, cerebral white matter, and arterial vascular territories from both non-PVEc and PVEc ASL data (28 features). Additionally, perfusion ratios between LICA and RICA, between LCWM and RCWM, and between GM and WM were from both non-PVEc and PVEc ASL data also included as features (6 features). Furthermore, similarity metrics (5 features) for registration were also incorporated. In total, 43 features were obtained for building the machine learning models.

SVMs with gaussian kernel were constructed for each individual feature to evaluate their respective capabilities in detecting outliers. Additionally, an SVM model with gaussian kernel was developed using the Python scikit-learn package [230] combining all metrics together to assess the overall effectiveness of QC. All SVM models were trained using 5-fold cross-validation to ensure robustness and prevent overfitting. To evaluate the quality of the classification, several metrics were computed from the confusion matrix, including the accuracy, FPR (false positive rate), sensitivity (true positive rate), and specificity (true negative rate).

In addition to these metrics, permutation feature importance was calculated for each feature. This method involved breaking the relationship between each feature and the model to assess how much the model performance decreased when that a feature was permuted. This approach allowed for a comparison of the performance of each metric individually and in combination, providing insights into which metrics or combinations thereof are most effective for identifying outliers in ASL data.

4.4 Results

4.4.1 Image Quality Metrics

The T1w images and ASL images were successfully processed in 474 subjects and the quality metrics were computed using ASL images. The distributions for each metric in the cohort are presented using boxplots: SNR, CNR, tSNR, and tCNR in Figure 4.4.1; non-PVEc CBF and SpCoV in Figure 4.4.2; PVEc CBF and SpCoV in Figure 4.4.3; the registration quality metrics in Figure 4.4.4.

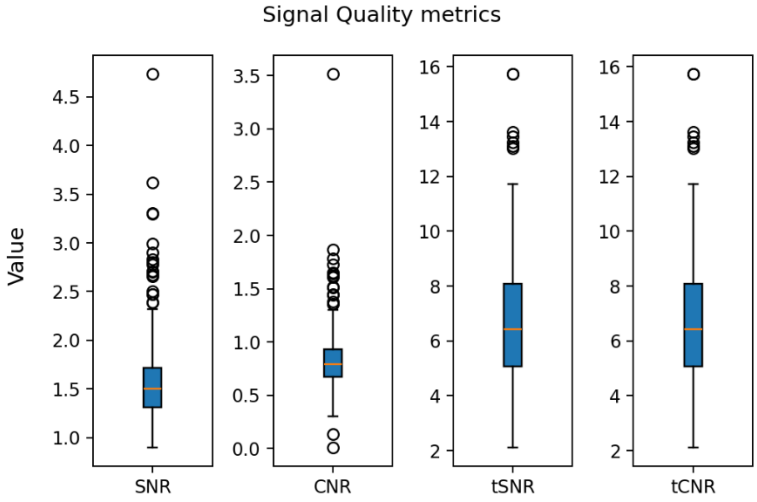
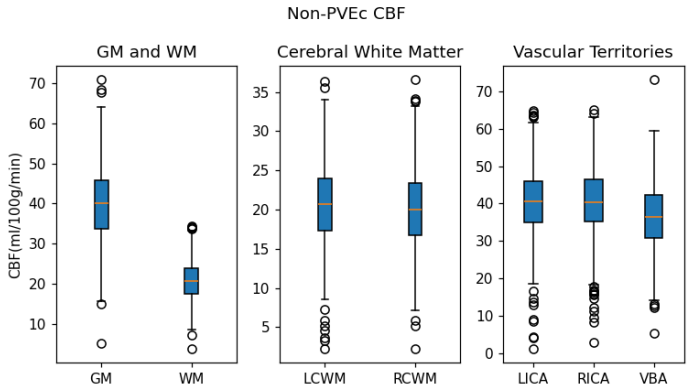


Figure 4.4.1 Box plots of distributions of signal quality metrics for all subjects (from left to right): SNR, CNR, tSNR, and tCNR.



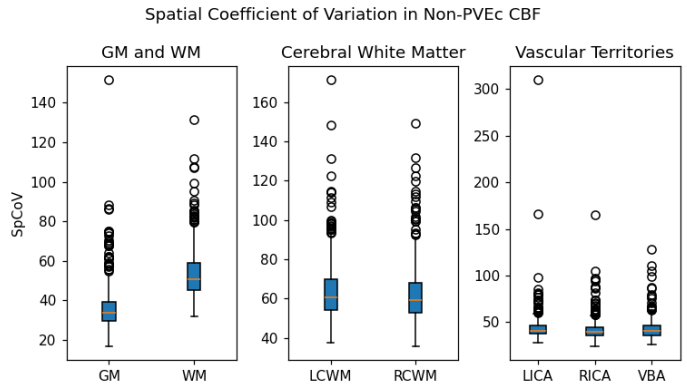


Figure 4.4.2 Box plots of distributions of non-PVEc CBF (Top) and SpCoV (bottom) in ROIs for all subjects. LCWM: left cerebral white matter; RCWM: right cerebral white matter; LICA: left internal carotid artery; RICA: left right internal carotid artery; VBA: vertebrobasilar artery.

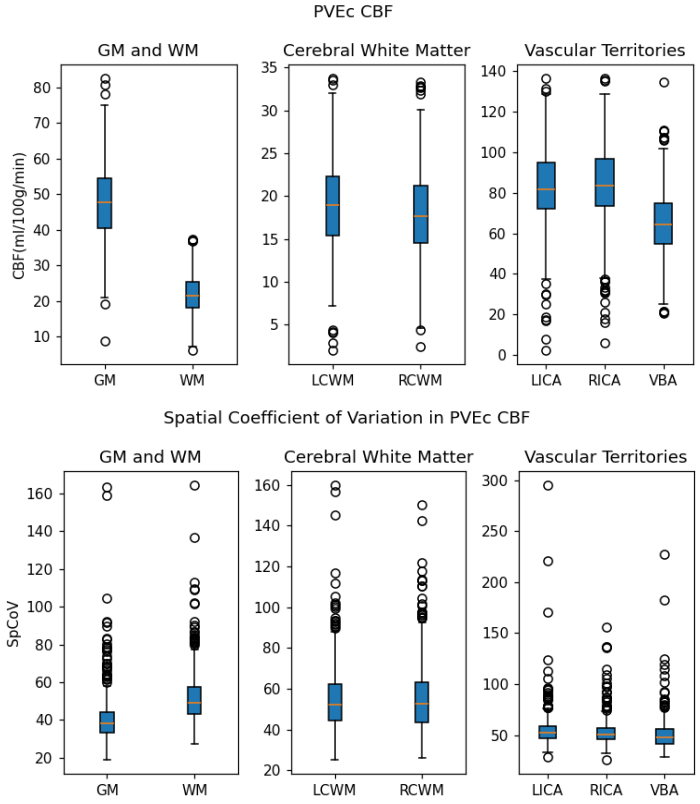


Figure 4.4.3 Box plots of distributions of PVEc CBF (Top) and SpCoV (bottom) in ROIs for all subjects. LCWM: left cerebral white matter; RCWM: right cerebral white matter; LICA: left internal carotid artery; RICA: left right internal carotid artery; VBA: vertebrobasilar artery.

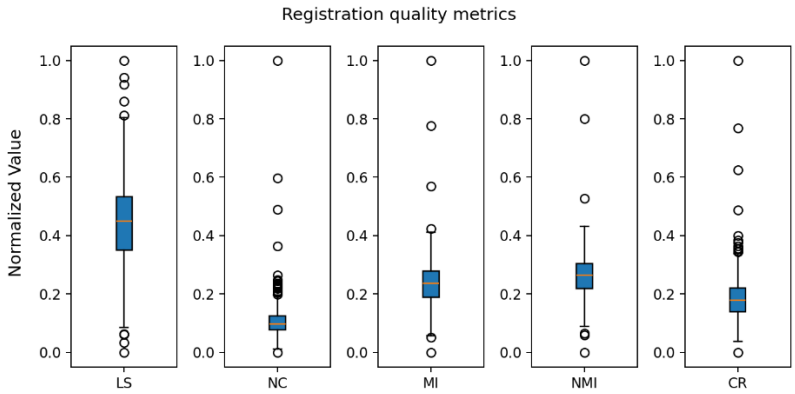


Figure 4.4.4 Box plots of distributions of normalized registration quality metrics over all subjects.

Least Square (LS), Normalized Correlation (NC), Mutual Information (MI), Normalized Mutual Information (NMI), and Correlation Ratio (CR).

4.4.2 Quality Control Report

The example results from group-level QC report are presented including signal quality metrics in Figure 4.4.5; non-PVEc and PVEc CBF in Figure 4.4.6; non-PVEc and PVEc SpCov in Figure 4.4.7; and registration quality metrics (Figure 4.4.8). Dropdown menus were provided to choose specific metric; the upper and lower thresholds can be set by sliders. The filtered subjects were displayed in the text area below.

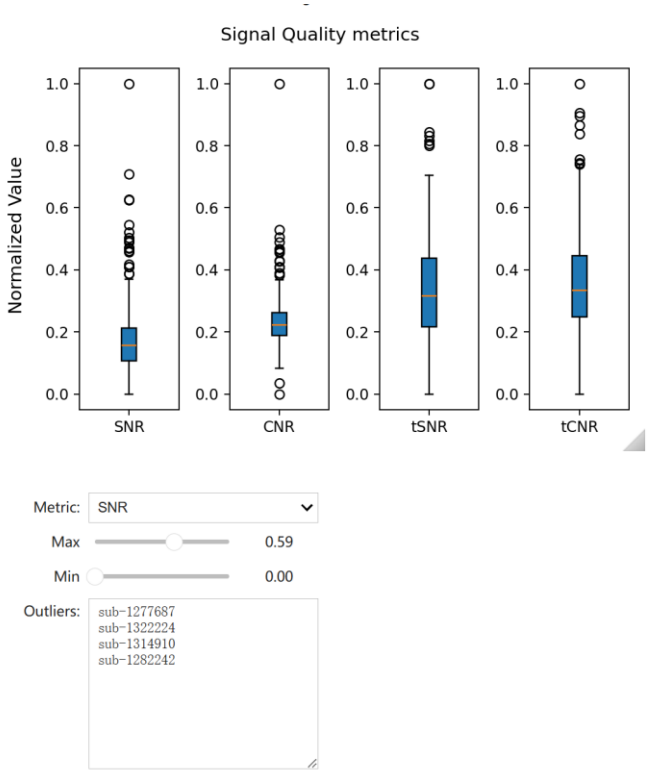


Figure 4.4.5 The example results from group-level report for normalized signal quality metrics: SNR, CNR, tSNR, and tCNR. The selected metric can be set with the metric dropdown menu. Upper and lower thresholds can be set by the Max and Min sliders, respectively, and thus the outliers can be filtered.

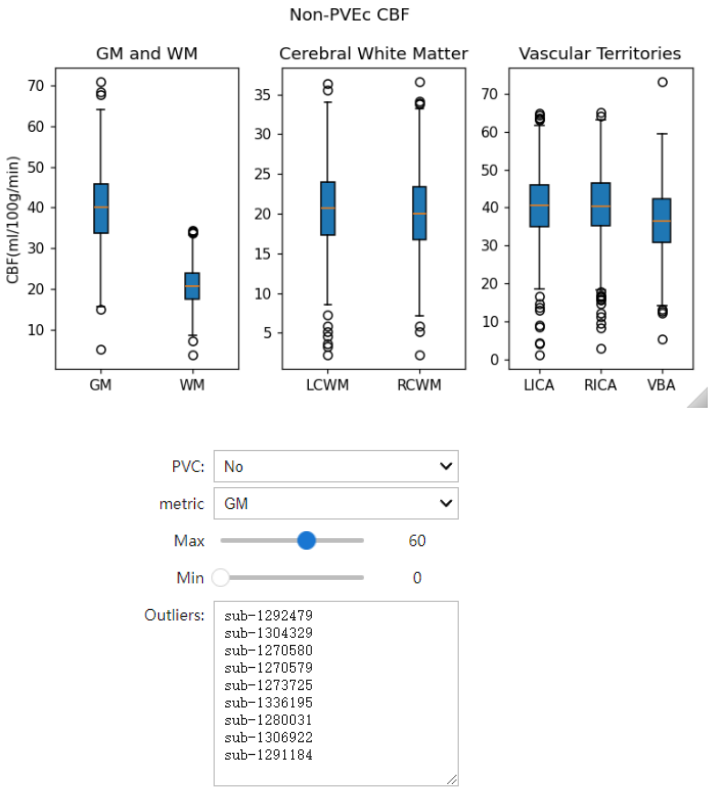


Figure 4.4.6 The example results from group-level report for CBF values on GM, WM, cerebral WM and arterial vascular territories. PVE or non-PVEc CBF maps can be set using the PVC dropdown menu, and the selected metric can be set with the metric dropdown menu. Upper and lower thresholds can be set by the Max and Min sliders, respectively, and thus the outliers can be filtered. Non-PVEc or PVEc results can be selected by setting the PVEc dropdown menu. LCWM: left cerebral white matter; RCWM: right cerebral white matter; LICA: left internal carotid artery; RICA: left right internal carotid artery; VBA: vertebrobasilar artery.

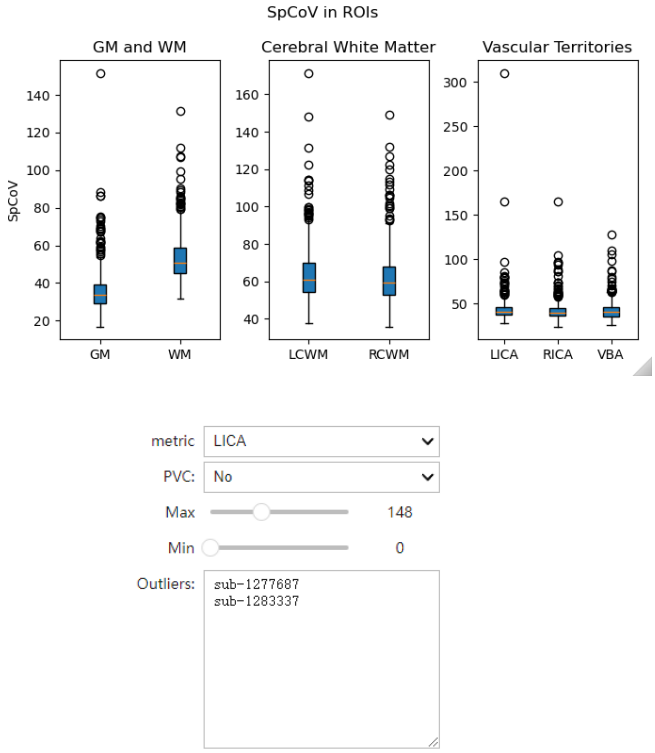
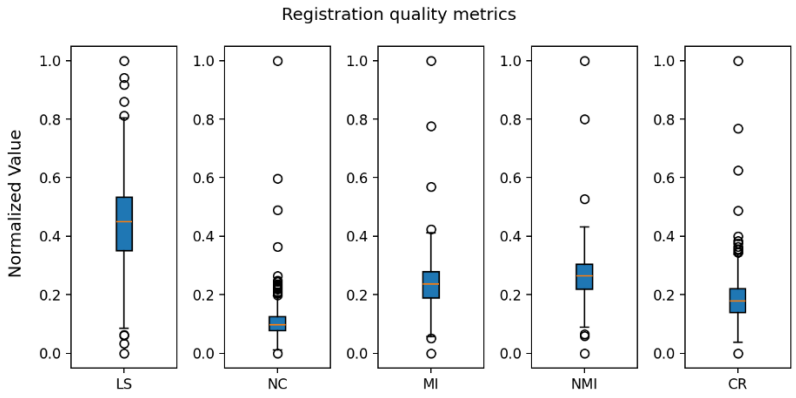


Figure 4.4.7 The example results from group-level report for spatial coefficient of variation (SpCov) values on GM, WM, cerebral WM and arterial vascular territories. PVE or non-PVEc CBF maps can be set using the PVC dropdown menu, and the selected metric can be set with the metric dropdown menu. Upper and lower thresholds can be set by the Max and Min sliders, respectively, and thus the outliers can be filtered. Non-PVEc or PVEc results can be selected by setting the PVEc dropdown menu. LCWM: left cerebral white matter; RCWM: right cerebral white matter; LICA: left internal carotid artery; RICA: left right internal carotid artery; VBA: vertebrobasilar artery.



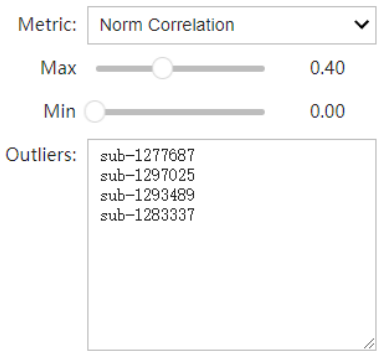
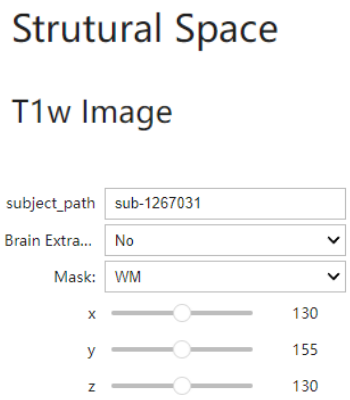


Figure 4.4.8 The example results from group-level report for normalized registration quality metrics. The selected metric can be set with the metric dropdown menu. Upper and lower thresholds can be set by the Max and Min sliders, respectively, and thus the outliers can be filtered. LS: Least Squares, NC:Normalized Correlation, MI: Mutual Information, NMI: Normalized Mutual Information, and CR: Correlation Ratio.

Individual QC report

The example results from individual-level QC report were presented including the structural space QC in Figure 4.4.9; the ASL native QC section for perfusion-weighted images before and after motion and distortion correction in Figure 4.4.10; CBF maps in Figure 4.4.11; regional analysis in Figure 4.4.12; and motion estimates in Figure 4.4.13; the standard space QC in Figure 4.4.14.



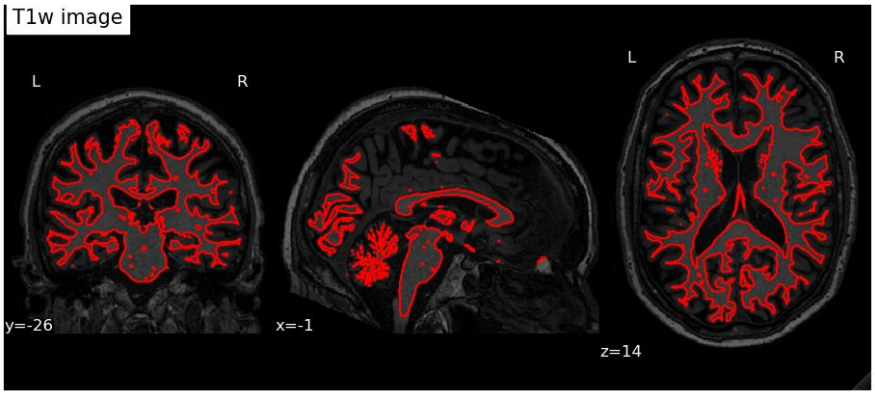


Figure 4.4.9 The example of the results from individual-level report for the structural space QC. Brain contour can be overlain on the structural image to check PVs segmentation.

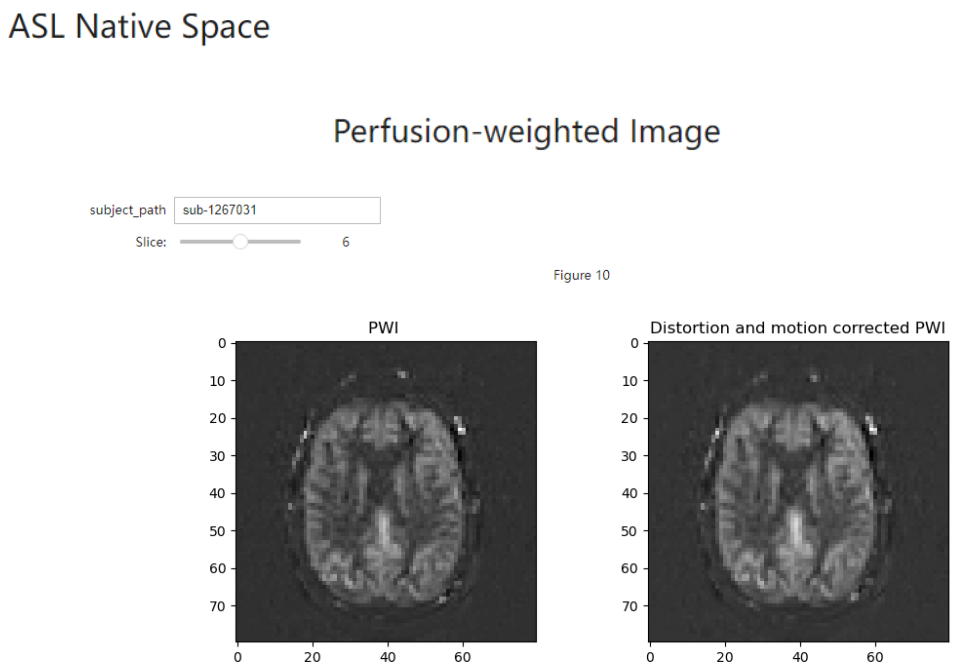


Figure 4.4.10 The example of the perfusion-weighted image. The left image is without any correction and the right is the PWI image after distortion correction and motion correction. The visualized slice position can be adjusted from the slider.

CBF map

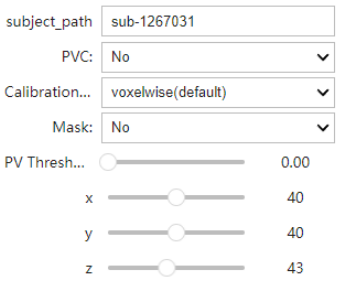


Figure 11

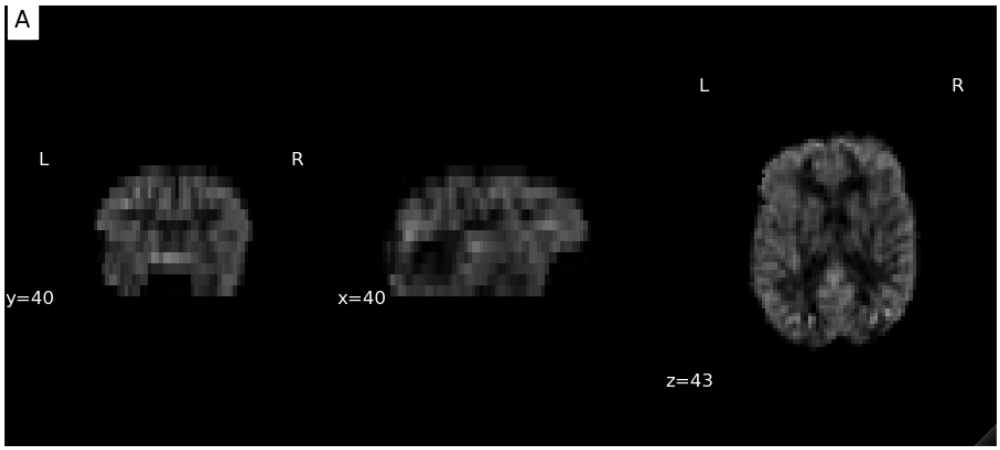


Figure 4.4.11 The example results from individual-level report for CBF maps in the native ASL space, featuring dropdown menus that allow users to customize the display by selecting non-PVEc or PVEc GM CBF maps, choosing different calibration methods (voxelwise, WM, or CSF as reference regions), and adjusting GM/WM mask overlays with threshold settings.

Region Analysis

subject_path	sub-1267031
PVC:	No
Calibration...	voxelwise(default)

	name	Nvoxels	Mean	Std	Median	IQR	Precision-weighted mean	I2	SpCov
0	GM	271	41.13	14.790	39.22	17.30	41.94	84.0	35.96
1	WM	1743	20.43	9.638	19.51	12.70	18.79	68.0	47.18
2	RCWM	1137	20.00	11.200	18.28	13.90	19.21	80.0	56.00
3	LCWM	1129	22.35	12.660	20.48	15.51	22.11	86.0	56.64
4	RICA	3793	43.49	14.290	43.23	19.96	42.69	88.0	32.86
5	LICA	3813	43.83	14.400	43.71	19.71	42.95	89.0	32.85
6	VBA	1354	39.76	15.120	38.56	17.61	40.64	93.0	38.03

	LICA	RICA	L/R
0	43.83	43.49	1.01

	LCWM	RCWM	L/R
0	22.35	20.0	1.12

	GM	WM	GM/WM
0	41.13	20.43	2.01

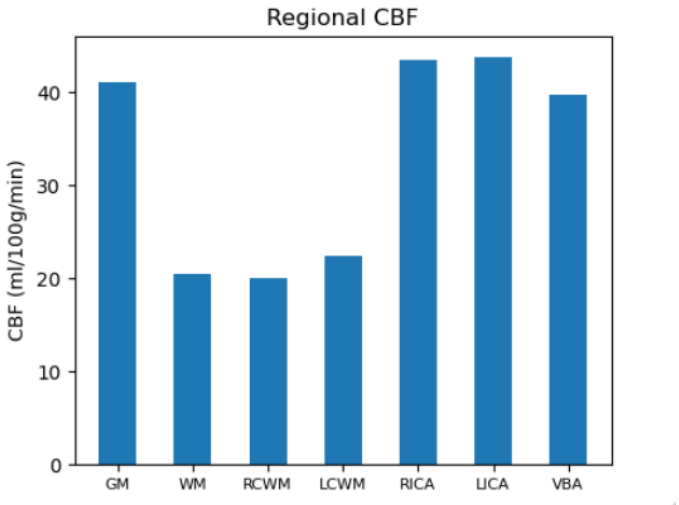


Figure 4.4.12 The example results from individual-level report for regional analysis presenting CBF values and SpCoV in global GM and WM, as well as in specific ROIs. It also includes ratios between LICA and RICA, between LCWM and RCWM, and between GM and WM, along with histograms of CBF values in these regions.

Motion estimates

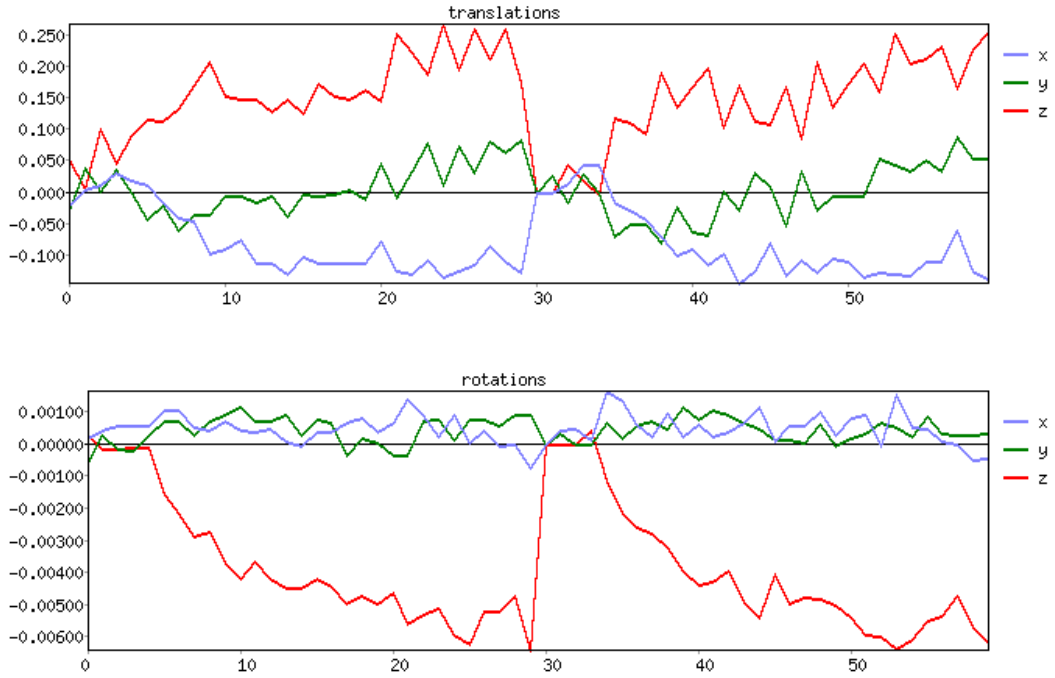


Figure 4.4.13 The example of the results from individual-level report for motion estimates in rotations (top) and translations (bottom) of the ASL image. The x axis contains 30 interleaved label and control images, and y axis is the shifted distance with the middle volume as the reference, and x,y,z represent the shifts in three dimensions.

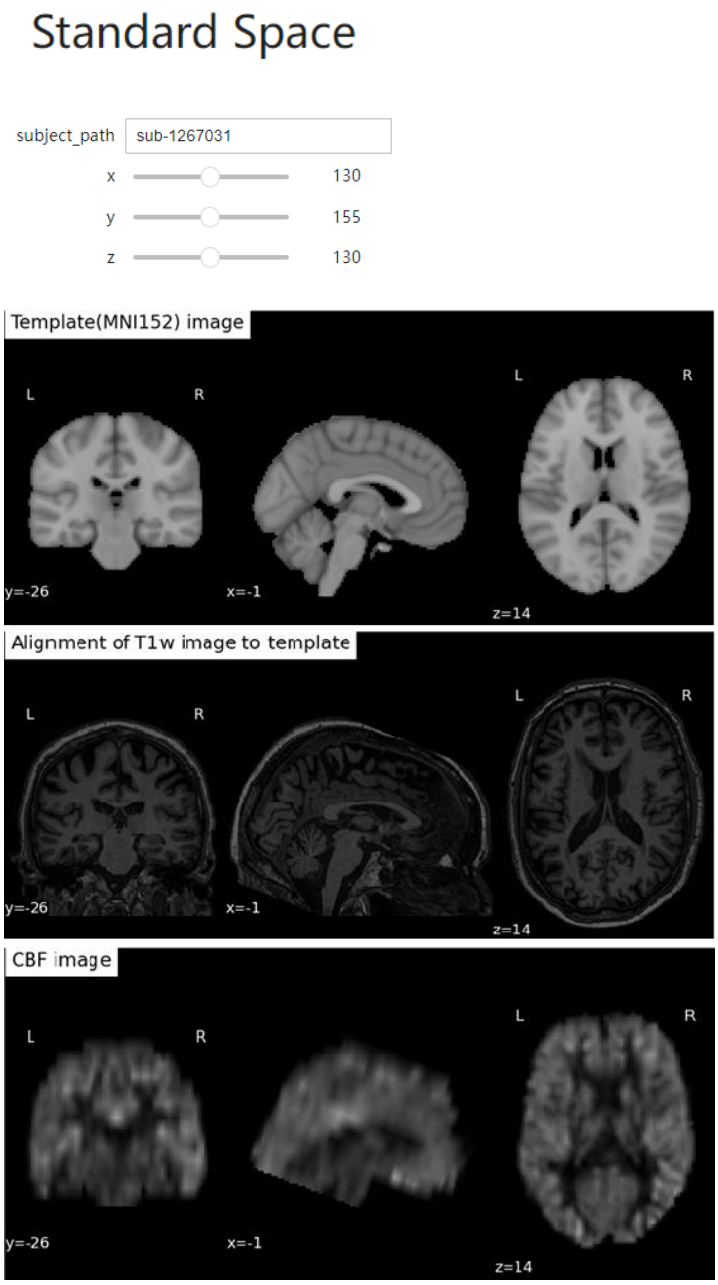


Figure 4.4.14 The example results from individual-level report for the template image, T1w image, and CBF image aligned to the standard template.

4.4.3 Quality Control Using Machine Learning

The classification results for each individual feature are presented in Table 4.4.1.

Table 4.4.1 Classification on test dataset using each quality metric as the individual feature.

Feature	TP	FN	FP	TN	Accuracy	Sensitivity	Specificity
NonPVC GM CBF	138	0	4	1	0.97	1	0.2
NonPVC GM SpCov	138	0	3	2	0.98	1	0.4
NonPVC WM CBF	137	1	2	3	0.98	0.99	0.6
NonPVC WM SpCov	138	0	4	1	0.97	1	0.2
NonPVC RCWM CBF	137	1	4	1	0.97	0.99	0.2
NonPVC RCWM SpCov	135	3	3	2	0.96	0.98	0.4
NonPVC LCWM CBF	138	0	3	2	0.98	1	0.4
NonPVC LCWM SpCov	136	2	2	3	0.97	0.99	0.6
NonPVC VBA CBF	138	0	3	2	0.98	1	0.4
NonPVC VBA SpCov	137	1	3	2	0.97	0.99	0.4
NonPVC RICA CBF	137	1	4	1	0.97	0.99	0.2
NonPVC RICA SpCov	138	0	1	4	0.99	1	0.8
NonPVC LICA CBF	138	0	3	2	0.98	1	0.4
NonPVC LICA SpCov	138	0	3	2	0.98	1	0.4
SNR	138	0	5	0	0.97	1	0
tSNR	138	0	5	0	0.97	1	0
CNR	138	0	5	0	0.97	1	0
tCNR	138	0	5	0	0.97	1	0
Least Square	138	0	3	2	0.98	1	0.4
Norm Correlation	136	2	2	3	0.97	0.99	0.6
Mutual Information	138	0	5	0	0.97	1	0
Norm Mutual Information	138	0	5	0	0.97	1	0
Correlation Ratio	136	2	2	3	0.97	0.99	0.6
NonPVC LICA/RICA	138	0	2	3	0.99	1	0.6
NonPVC LCWM/RCWM	138	0	2	3	0.99	1	0.6
NonPVC GM/WM	138	0	5	0	0.97	1	0
PVC GM CBF	138	0	4	1	0.97	1	0.2
PVC GM SpCov	136	2	4	1	0.96	0.99	0.2
PVC WM CBF	137	1	2	3	0.98	0.99	0.6
PVC WM SpCov	138	0	4	1	0.97	1	0.2
PVC RCWM CBF	137	1	4	1	0.97	0.99	0.2
PVC RCWM SpCov	138	0	5	0	0.97	1	0
PVC LCWM CBF	138	0	3	2	0.98	1	0.4
PVC LCWM SpCov	138	0	5	0	0.97	1	0
PVC VBA CBF	138	0	4	1	0.97	1	0.2
PVC VBA SpCov	137	1	3	2	0.97	0.99	0.4
PVC RICA CBF	137	1	4	1	0.97	0.99	0.2
PVC RICA SpCov	137	1	1	4	0.99	0.99	0.8
PVC LICA CBF	138	0	3	2	0.98	1	0.4
PVC LICA SpCov	138	0	5	0	0.97	1	0
PVC LICA/RICA	138	0	2	3	0.99	1	0.6
PVC LCWM/RCWM	138	0	2	3	0.99	1	0.6
PVC GM/WM	138	0	5	0	0.97	1	0

TP: True Positive; FP: False Positive; TN: True Negative; FN: False Negative; Accuracy: T/(T+N); Sensitivity: TP/(TP+FN); Specificity: TN/(TN+FP); FPR: False Positive rate (FP/(FP+TN)).

The classification using all features in test dataset is shown in Figure 4.4.15. The accuracy, FPR, sensitivity, and specificity were 0.97, 0.2, 0.99, and 0.8 respectively. According to the classification results, the examples of detected outliers (True-Negative), not detected outliers (False-Negative), and mis-classified normal images (False-Positive) are presented in Figure 4.4.16, Figure 4.4.17, and Figure 4.4.18,

respectively. The permutation feature importances, shown in Figure 4.4.19, is a model-agnostic technique used to assess the importance of individual features in thea trained model by comparoncomparisons with the re-trained the model with excluding each feature..

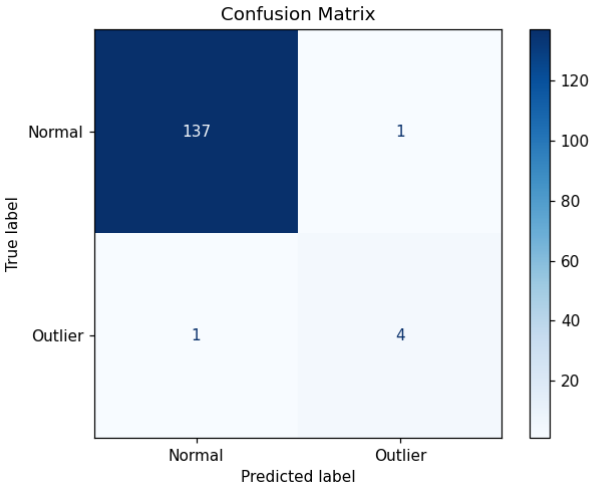


Figure 4.4.15 Confusion matrix of classification by SVM in test data.

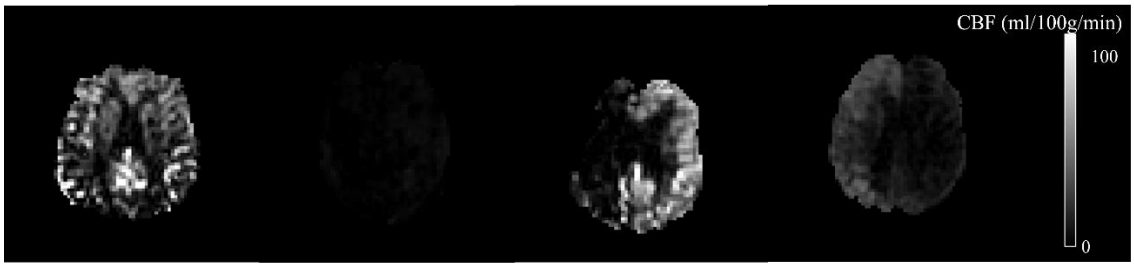


Figure 4.4.16 Detected outliers in the test dataset. The artifact types are (from left to right): delayed arrival, poor signal, poor labelling, poor labelling.

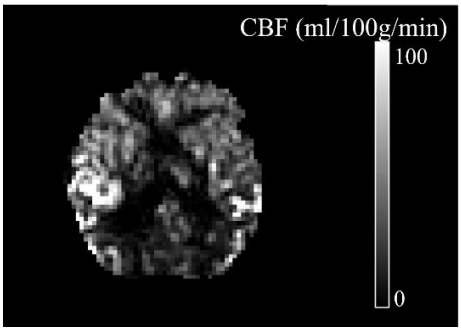


Figure 4.4.17 The outlier (delayed arrival) was not detected in the test dataset.

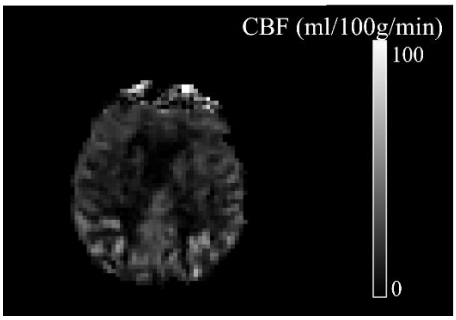


Figure 4.4.18 The normal CBF map classified as outlier in the test dataset. In the image, hyper-perfusion voxels are observed in the anterior brain, potentially due to imperfect distortion correction. This could be regarded as one outlier; however, this was not among the predefined labelled artifact types.

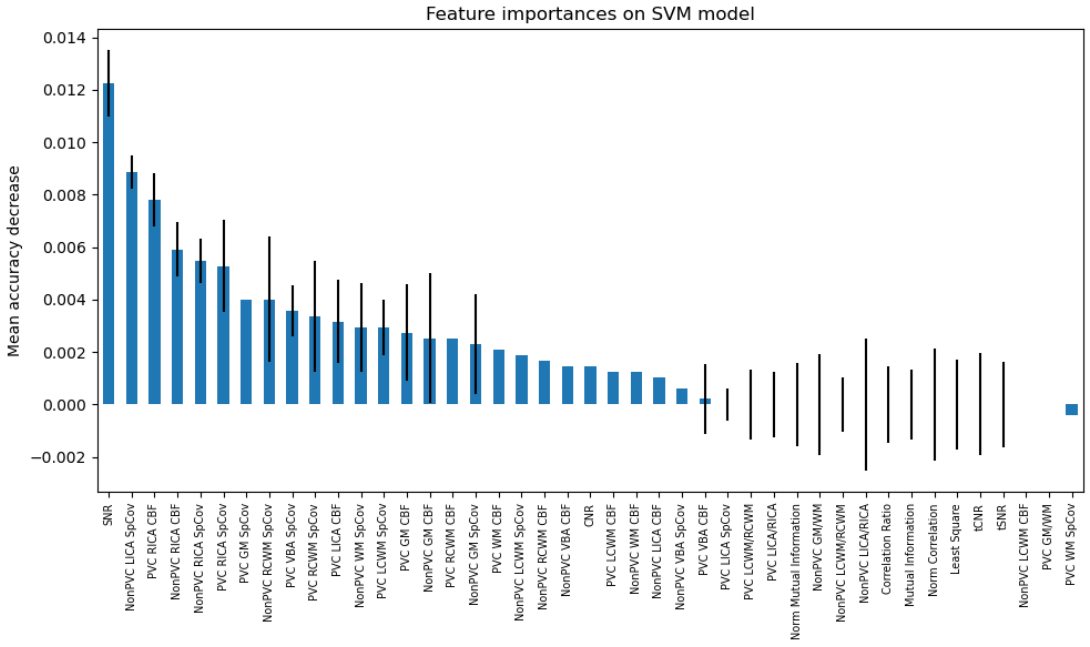


Figure 4.4.19 The permutation feature importances in the SVM model. Mean accuracy decrease: the decreased accuracy after excluding each feature.

4.5 Discussion

This work aimed to develop standardized QC protocols for ASL MRI with OXASL, which was achieved through three key steps. First, image metrics specific to ASL MRI were investigated to address the unique challenges of measuring perfusion through blood water labelling. Second, QC reports were generated to help identify outliers and assist in the screening of ASL images processed by OXASL. Finally, machine learning models were developed to evaluate the reliability and efficiency of these ASL quality metrics on a large ASL dataset, with the goal of providing effective and automated QC for ASL data.

The ASL quality metrics included signal quality metrics, prefusion quality metrics, and registration quality metrics. The signal quality metrics, providing intrinsic characteristics of ASL acquisitions before any processing, are valuable for initial examination, as indicated by their high importance when considering all quality metrics. However, relying solely on signal-related metrics to judge the quality of ASL images is insufficient. This is evidenced by the poor performance of classifiers using these metrics (no outliers were detected), as they may not capture all issues affecting the images. For example, specific artifacts like signal loss might not significantly impact SNR or CNR but could still cause problems in perfusion quantification. To address this, perfusion-related metrics, including CBF and SpCoV within ROIs (such as GM, WM, cerebral WM, and arterial vascular territories), were computed on non-PVEc and PVEc ASL data to better account for the quality of ASL data quantification. Notably, RICA SpCoV from non-PVEc and PVEc data were particularly effective, identifying the most outliers (4 out of 5), while other metrics detected 3 or fewer outliers. Although these classifiers showed abilities to detect outliers, they were generally limited in scope, often identifying only a small number of outliers and exhibiting a high false positive rate (FPR), meaning that many outliers were missed. In addition, similarity metrics (such as LS, NC, and CR) for registration were able to detect some outliers, even though the labels were predefined for native ASL data rather than based on the success or failure of registration. However, mutual information or normalized mutual information did not demonstrate the same effectiveness. This suggests that while registration metrics can contribute to outlier detection, not all metrics are equally reliable for this purpose.

The group-level QC report was designed to filter extreme values of quality metrics that might indicate outliers. Given the difficulty of establishing absolute thresholds due to variations in study requirements, scanner types, and imaging protocols, users can define their own thresholds to identify potential outliers. Once these outliers are flagged, suspicious images can be further examined using the individual-level QC report. By utilizing both group-level and individual-level QC reports, the time required for QC can be significantly reduced. The individual-level reports focus on displaying only key images and quality metrics, rather than a large number of general quality images. This approach ensures the completeness of QC while being time-efficient, as it avoids redundant information. Additionally, since the QC reports are

generated post-processing in OXASL, they do not interfere with the processing itself, making this approach convenient and adaptable for application to other large ASL datasets.

The SVM model trained with all 43 ASL quality metrics demonstrated strong performance in outlier detection, even with the limited number of outliers in the test dataset. For QC purposes, it is generally acceptable to misclassify some normal images as outliers, as this might slightly increase the number of images requiring further examination but ensures that no true outliers are missed. This approach may result in high specificity but a low false positive rate. While precise classification is ideal, achieving it is challenging due to the extremely unbalanced nature of the data. The permutation feature importances for the SVM model using all features were provided, revealing that some features, such as SNR, PVEc RICA CBF, CNR, and non-PVEc LICA SpCoV, had higher importance. These features may be more sensitive to outliers, indicating they are closer to the decision boundary. However, features with lower or zero importance should not be considered useless; instead, they may simply be further from the decision boundary in this specific model.

There are some limitations to this study. The data were originally processed using BASIL FSL with CSF as the reference region for calibration, on which raters categorized all subjects into five categories. In this study, however, the data were processed using OXASL with voxel-wise calibration. Discrepancies in processed images due to different calibration methods may lead to differences in data labelling, such as the normal image misclassified as a outlier with imperfect distortion correction shown in Figure 4.4.18. Additionally, the trained machine learning classifier still needs to be validated on other ASL datasets processed by OXASL, as factors like scanner types and acquisition protocols may influence outlier identification. Furthermore, the quality metrics generated in this study are specific to ASL MRI data, particularly single-PLD data. In the future, quality metrics for multi-PLD ASL data will be needed, as there may be slight differences in their application. Furthermore, IQMs for general image QC may also be included to enhance the classifier performance. The QC reports were developed using Jupyter Notebook, allowing for interactive adjustments by users. However, this setup requires all files to be available on the development device, which is not ideal for processes that typically occur on high-performance clusters or servers. Future work will focus on

implementing these functions in more accessible formats, such as HTML, which can be easily shared and downloaded.

4.6 Conclusion

In this chapter, this study used 43 features specific to ASL were used for QC to establish standardized QC protocols for ASL, a gap in the current literature. Machine learning was employed to assess the effectiveness of these protocols and enable automatic QC for CBF maps, with the model proving robust against most artifacts in the dataset. QC reports were generated to assist with the manual review of ASL data.

While this machine learning approach demonstrated high specificity, it still depends on manually labelled data for training. Despite the limited number of outliers in the unbalanced dataset, the model was able to maintain strong classification performance. In the next chapter, deep learning techniques will be explored for quality control purposes, with the aim of improving outlier detection in ASL data. This shift to deep learning is anticipated to enhance both accuracy and automation in the quality control process, building upon the machine learning foundation established in this chapter.

5 Quality Control Using Deep Learning in ASL MRI Images

5.1 Introduction

ASL MRI provides the only truly non-invasive method to measure perfusion by magnetically labelling arterial blood water as an endogenous tracer [54] . However, ASL MRI, like other MRI techniques, is susceptible to artifacts, such as motion, distortion and poor labelling efficiency, which can compromise image quality and impair accurate CBF quantification [5] ,[82] . Therefore, the quality of ASL images needs to be assessed and ensured, before proceeding with any further statistical analysis. Quality Control (QC) is the process to identify and exclude outliers for MRI data which is conventionally conducted by experts who visually examine individual images. While this manual approach is necessary, it is time-consuming, subjective, and often impractical for large datasets [163] ,[171] .

Previous studies [9] ,[83] have increasingly employed machine learning to automate QC processes for MRI data. Machine learning models utilize Image Quality Metrics (IQMs) derived from MRI data to assess image quality, enabling them to identify potential issues based on predefined criteria. However, this approach faces several challenges. First, the definition and selection of IQMs can greatly influence the performance of machine learning models. The choice of metrics may not fully capture all relevant aspects of image quality, necessitating careful evaluation by experts to ensure that the selected IQMs are both comprehensive and appropriate for the task at hand. Second, supervised classification models still rely on labelled data, requiring expert annotations to guide model training—a process that remains subjective. In addition, acquiring labelled data for all types of potential anomalies in MRI data is impractical, which increases the risk of misclassifying unseen outliers. Furthermore, machine learning methods often struggle with unbalanced data, a common issue in QC tasks where the number of poor-quality images is typically much smaller than the number of high-quality images. This imbalance can result in models that are biased toward the majority class, reducing their effectiveness in detecting rare but critical anomalies.

Alternatively, deep learning offers a promising approach for QC in MRI, with QC in deep learning often referred to as anomaly detection. Unlike traditional machine learning, deep learning can automatically learn complex patterns and features directly from the image data, reducing the dependency on manually engineered IQMs and labels. This capability allows deep learning models to detect subtle anomalies that might be missed by both manual inspection and traditional machine learning approaches. Supervised deep learning approaches often utilize CNN-based networks as feature extractors, followed by one-class classifiers. However, they share the same limitation as supervised machine learning models in that they require manually labelled data for training.

Instead, semi-supervised approaches [232] ,[233] have emerged as a more plausible solution. These methods focus on identifying deviations from established norms, rather than relying on extensive labelled data. The initial step in this process involves developing a robust baseline of normal MRI scans, which serves as a reference point for detecting anomalies. This baseline is created by analysing a large dataset of MRI images that have been verified to be free of defects. Deep learning algorithms are employed to learn the typical patterns and features of these high-quality MRI images, establishing a standard against which new, unseen images can be compared. Once this baseline is established, the anomaly detection system can evaluate unseen MRI data by identifying deviations from the norm. If an image aligns with the patterns learned during training, the reconstruction error will be low, indicating a normal scan. Any abnormality that suggests potential anomalies are flagged for further inspection by comparing the reconstructed images against the established baseline. This semi-supervised approach thus enables the detection of anomalies without the need for extensive labelled datasets.

Building on this foundation, two prominent deep learning models Variational Autoencoders (VAEs) [119] and Generative Adversarial Networks (GANs) [120] have been effectively utilized for anomaly detection in MRI data. VAEs are generative models that learn to encode MRI images into a lower-dimensional latent space, capturing the essential features of the data, and then decode them back into the original image space. GANs, on the other hand, consist of two competing neural networks—a generator and a discriminator. The generator creates synthetic MRI images, while the discriminator attempts to distinguish between real and generated

images. The competition between the two networks enables the GAN to improve the generation of realistic images, and when used for anomaly detection, it can identify discrepancies between the generated and real data to flag potential outliers.

In this chapter, we employed a VAE-GAN model, a deep learning technique, to perform QC for ASL data using a deviation-based method allowing for effective detection of outliers in ASL data.

5.2 Theory

5.2.1 Variational Autoencoders

Autoencoders [244] are a type of artificial neural network designed to learn efficient data representations, typically for dimensionality reduction or feature learning. The conventional architecture of autoencoders is presented in Figure 5.2.1. They consist of two main components: an encoder, which compresses the input data into a lower-dimensional latent space, and a decoder, which reconstructs the original data from this compressed representation. The primary goal of an autoencoder is to minimize the reconstruction error, ensuring that the data compressed in the latent space retains enough information to accurately reconstruct the original input. A limitation of traditional autoencoders is that they do not explicitly model the distribution of the latent space, which can lead to poor generalization, especially when generating new data.

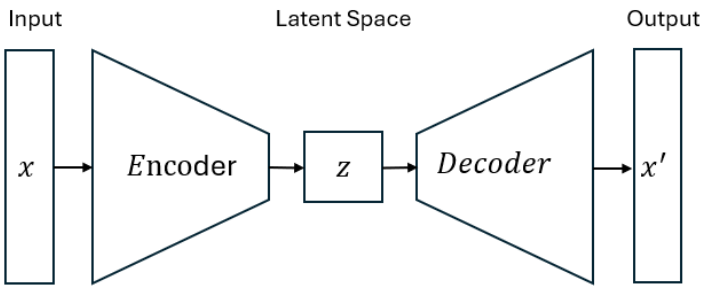


Figure 5.2.1 The architecture of an autoencoder.

It consists of an encoder, which compresses input data into a lower-dimensional latent representation, and a decoder, which reconstructs the original input from this representation.

Variational Autoencoders (VAEs) [119] build on the foundation of autoencoders but introduce a probabilistic framework. Instead of mapping input data to a single point in the latent space, VAEs learn to map it to a distribution, typically modelled as a Gaussian. During training, the encoder learns the mean and variance of this distribution for each input. The decoder then reconstructs the input data by sampling from this latent distribution.

Let z represent the latent variable and x the observed data. The probability of z in the latent space, given the observation x, can be expressed using Bayes' rule as:

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} \qquad \textit{Equation 5.2.1}$$

where $p(x) = \int p(x|z)p(z)dx$ is the marginal likelihood. The marginal likelihood p(x) is intractable in most cases, however, the value can be estimated using variational inference by assuming that $p(z|x)$ is approximated by another distribution, denoted as $q(z|x)$, which has a tractable distribution. This can be achieved by minimizing the Kullback–Leibler (KL) divergency, a measure of difference between two probability distributions. Furthermore, the minimization KL divergence can be converted to maximizing the following:

$$\mathcal{L}_{VAE} = E_{q(z|x)} \log p(x|z) - KL(q(z|x)||p(z|x)) \quad \textit{Equation 5.2.2}$$

The first term to penalize reconstruction error and the second term to ensure that the learned distribution q is similar to the prior distribution p. Assuming a Gaussian likelihood, the loss function used in the VAE structure can be simplified as:

$$\mathcal{L}_{VAE} = \mathcal{L}_{Rec} + \sum_j KL(q_j(z|x)||p(z)) \quad \textit{Equation 5.2.3}$$

Notably, optimizing the KL divergence term requires a delicate balance with the reconstruction loss in practice.

5.2.2 Generative Adversarial Networks

Generative Adversarial Networks (GANs) [120] are a class of deep learning models comprising two neural networks, a generator and a discriminator, that compete against each other in a zero-sum game. The generator learns to produce data that is increasingly realistic over time, while the discriminator learns to better identify fake

data. This adversarial process continues until the generator's output becomes indistinguishable from real data, indicating that it has successfully learned the underlying distribution. The architecture of a GAN is shown in Figure 5.2.2.

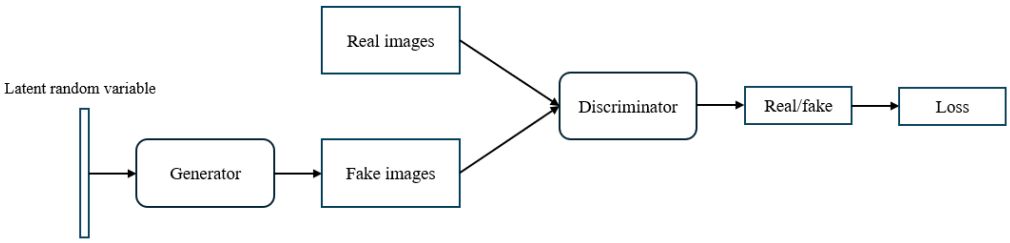


Figure 5.2.2 The architecture of a GAN, which consists of two neural networks: a generator and a discriminator.

The generator network, Gen(z), maps a latent random variable z to the data space, while the discriminator network assigns a probability y = Dis(x) ∈ [0,1], where y represents the likelihood that x is a real training sample, and 1 − y is the probability that x was generated by the model through x = Gen(z), with z ∼ p(z). The objective of the GAN is to find the optimal binary classifier that can best differentiate between real and generated data while simultaneously encouraging Gen to learn and fit the true data distribution. The loss function of GANs is designed to optimize this min-max problem using binary cross-entropy in Equation 5.2.4.

$$\mathcal{L}_{GAN} = \min_{Gen} \max_{Dis} log\big(Dis(x)\big) + log\,(1 - Dis(Gen(z)))\quad \textit{Equation 5.2.4}$$

On the one hand, discriminator aims at predicting $D$is(x) = 1 for real data samples and $D$is(Gen(z)) = 0 for fake samples. On the other hand, the GAN learns how to fool D by finding Gen which is optimized on hampering the second term in Equation 5.2.4.

5.2.3 VAE-GAN

In models such as [119] [234] , the selection of an appropriate similarity metric is crucial, as it forms the core of the training signal through the reconstruction error objective. Typically, element-wise measures like the squared error are employed by default. Although these metrics are straightforward, they are not particularly well-suited for image data, as they fail to account for the nuances of human visual

perception. For instance, a minor image translation can result in a significant pixel-wise error, despite being nearly imperceptible to the human eye. Consequently, it is preferable to measure image similarity using a higher-level, sufficiently invariant representation of the images.

Instead of manually engineering a metric to address the limitations of element-wise measures, it is more effective to design a task-specific function. The VAE-GAN model was developed to learn such a similarity measure by jointly training a VAE and a GAN [235] , utilizing the GAN discriminator to assess sample similarity. The VAE-GAN model architecture is illustrated in Figure 5.2.3, where the VAE-GAN model combines the VAE with the GAN by merging the decoder and generator into a single unit and training them simultaneously. In this model, the conventional element-wise reconstruction metric is replaced by a feature-wise metric derived from the discriminator.

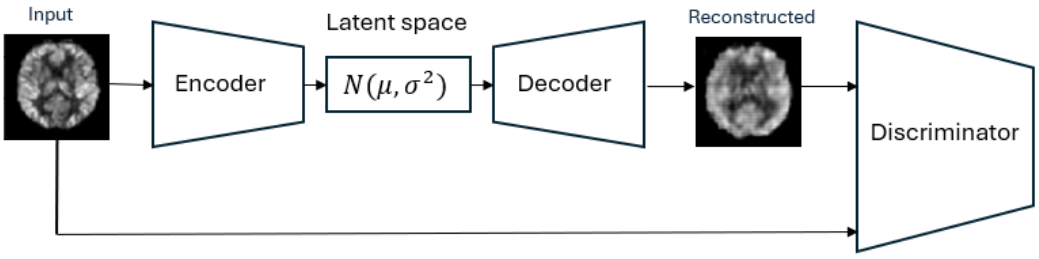


Figure 5.2.3 The architecture of the VAE-GAN model in this study. Decoder from VAE and the generator from GAN were combined into one. Image generation is achieved by sampling the latent space of the VAE, followed by decoding through decoder.

In VAE-GAN, the VAE reconstruction (expected log likelihood) error term from Equation 5.2.3 was replaced by a reconstruction error expressed in the GAN discriminator. This is achieved by using the hidden representation of the l-th layer of the discriminator, denoted as $Dis_l(x)$. Then Gaussian observation model was used to compute its mean $Dis_l(\tilde{x})$ and identity covariance:

$$p(Dis_l(x)|z) = N(Dis(x)|Dis_l(\tilde{x}), I) \qquad \textit{Equation 5.2.5}$$

Thus, the reconstruction error in VAE can be replaced with:

$$\mathcal{L}_{dis\_l} = -E_{q(z|x)}(log\ \ p(Dis_l(x)|z)) \qquad \textit{Equation 5.2.6}$$

The loss function of VAE-GAN incorporates three parts (depicted in Figure 5.2.4):

$$\mathcal{L}_{VAE-GAN} = \mathcal{L}_{dis\_l} + \mathcal{L}_{KL} + \mathcal{L}_{GAN} \qquad \textit{Equation 5.2.7}$$

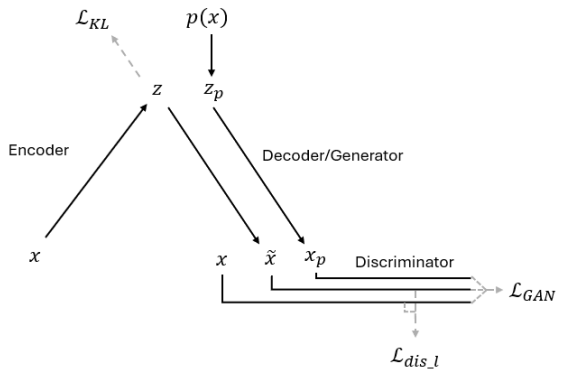


Figure 5.2.4 Loss functions in VAE-GAN including the discriminator loss, GAN loss, and KL loss.

5.2.4 Kernel Density Estimation

Kernel Density Estimation (KDE) [236] is a non-parametric method used to estimate the probability density function (PDF) of a random variable [236] . Unlike parametric methods, which assume a specific distribution for the data (e.g., normal distribution), KDE makes no such assumptions and can therefore adapt to the underlying structure of the data more flexibly.

The basic idea behind KDE is to estimate the probability density function by placing a kernel (a smooth, symmetric function) at each data point. The contribution of each data point to the overall density estimate is determined by the kernel function. The KDE then sums these contributions across all data points to produce a smooth estimate of the density function.

Given a set of data points $x_1, x_1, \ldots, x_n$, the kernel density estimation $f(x)$ at a point x is defined as:

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \qquad \textit{Equation 5.2.8}$$

Where n is the number of data points; h is the bandwidth, a smoothing parameter that controls the width of the kernel and, therefore, the smoothness of the density estimate; K is the kernel function, which is usually a symmetric, non-negative function that integrates to one. Common choices for the kernel function include the Gaussian (normal) kernel, Epanechnikov kernel, and uniform kernel.

KDE is effective in identifying multiple modes or peaks in distributions without making strong assumptions, enabling it to adapt to the true underlying shape of the

data. This is particularly useful in cases where the data distributions do not follow simple or well-known forms (such as Gaussian), making KDE a sensible approach for estimating the distribution [236] .

5.3 Methods

5.3.1 Dataset

This study used the same ASL dataset [166] as in Chapter 3 and 4. Like in chapter 4, all available subjects were retained for analysis. The non-PVEc CBF maps in the standard space, processed by the volumetric ASL pipeline in Chapter 3, were employed for quality control in this chapter, encompassing a total of 474 subjects.

The CBF maps were manually examined and labelled in the native space by two experts [163] , as in chapter 4. A total of 30 subjects were marked as outliers based on the types of artifacts present (see Figure 5.3.1 for examples). The artifact types included poor labelling efficiency (27 subjects), low contrast (1 subject), motion (7 subjects), poor signal (1 subject), and delayed arrival (6 subjects), with some outliers exhibiting multiple artifact types.

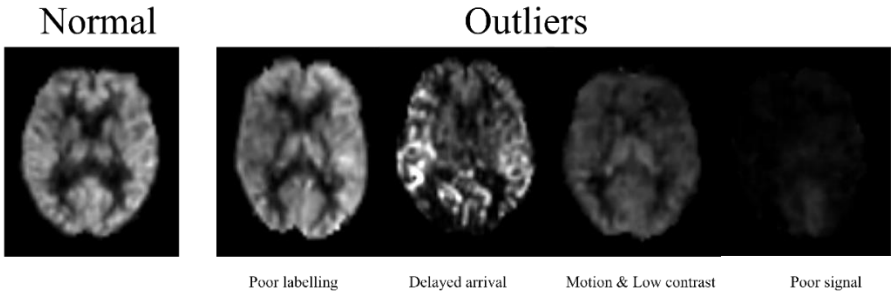


Figure 5.3.1 Normal (left) and outliers (right four images) examples in the standard space in the dataset.

Notably, the native CBF maps were registered to the MNI152 space [70] via intermediate structural images, and the CBF maps in standard space were then utilized in subsequent processing. The CBF images were resampled to a 64x64x64

resolution, with 2D images extracted from the central axial slice, to maintain computational efficiency and avoid the undying increase in data dimension from 2D to 3D. This resolution strikes a balance between computational efficiency and anatomical relevance, considering the relatively fewer layers of the models. Subsequently, the central axial slices from the CBF maps were randomly divided into a training set, consisting of 402 normal images, and a test set, comprising 42 normal images along with 30 outliers.

5.3.2 VAE-GAN Architecture

The architecture of the VAE-GAN network (shown in Figure 5.2.3) follows that with the details in Table 5.3.1.

Table 5.3.1 Architectures of encoder, decoder and discriminator for the VAE-GAN.

Encoder	Decoder	Discriminator
3x3 32 Conv ↓ BNorm, ReLU	8x8x32 FC, BNorm, ReLU	3x3 32 Conv ↓ BNorm, ReLU
3x3 64 Conv ↓ BNorm, ReLU	3x3 128 Conv ↑ BNorm, ReLU	3x3 64 Conv ↓ BNorm, ReLU
3x3 128 Conv ↓ BNorm, ReLU	3x3 64 Conv ↑ BNorm, ReLU	3x3 128 Conv ↓ BNorm, ReLU
2 FC, BNorm, ReLU	3x3 32 Conv ↑ BNorm, ReLU	512 FC, BNorm, ReLU
	3x3 1 Conv, tanh	1 FC, sigmoid

BNorm: Batch normalization; ReLU: Leaky ReLU; ↓ and ↑:downsamling and upsampling; Conv: convolutional layer with kernel size and the number of output channels listed ahead of it; FC: fully-connected layer with the number of output channels listed ahead of it. Stride is 2 in all convolutional layers.

5.3.3 Detect Outliers Using VAE-GAN

The process of outlier detection in this study is shown in Figure 5.3.2. A VAE-GAN model was used to capture the patterns of normal CBF maps. A significantly compressed latent space was employed to ensure that the latent space distribution remained tightly clustered, allowing the reconstructed images of the sampled latent space closely resembled typical normal data patterns This approach effectively accounted for the inherent variability among normal CBF images while maintaining

a focus on common structural features, thereby improving its ability to differentiate between normal and anomalous data.

During training, the difference scores between the input CBF images and the reconstructed images from the VAE-GAN were computed using metrics such as mean squared error (MSE), cosine similarity (CS), or both [237]. KDE was then applied to learn and model the distribution of these difference scores. In the testing phase, the same process was followed for the test data to calculate the difference scores, after which KDE was used to estimate the position of these images within the established distribution.

By modelling the distribution of the difference scores with KDE, thresholds were set using predefined criteria, allowing for the identification and filtering of outliers that exhibited significant deviations from the norm.

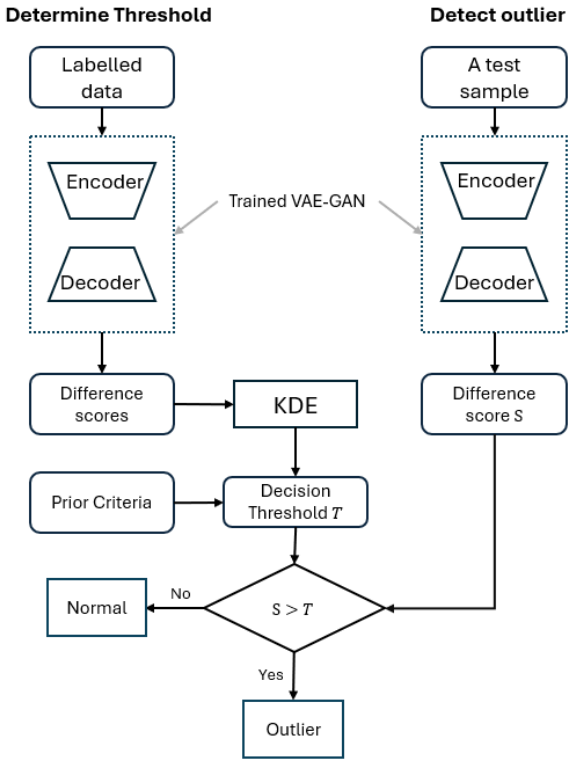


Figure 5.3.2 The workflow of QC using VAE-GAN. In training, KDE is used to learn the distribution of difference scores from VAE-GAN using normal images as input, and the threshold needs to be determined. In test, the quality of input image is judged by comparing its difference score to the determined threshold.

During training, all 402 normal images were fed into the VAE-GAN model to learn the distribution of normal data. The Adam optimizer was utilized, with learning rates

set at 0.0001 for the encoder, 0.001 for the decoder, and 0.0001 for the discriminator. The model was trained for up to 1000 epochs using the normal images. The third layer of the discriminator was used to calculate the feature-wise error between input images and reconstructed images to replace the element-wise error. A relatively low latent space size of 2 was chosen to avoid sparsity in the latent space. The difference scores consisted of the cosine similarity, mean squared error and their combination, which were computed between the input data and the reconstructed data. Then, KDE was used to estimate probability density function of the difference scores on the normal data with gaussian kernel and bandwidth of 0.2.

In testing, only the encoder and decoder are used like a traditional VAE. The test dataset comprising 42 normal images along with 30 outliers was input to the VAE-GAN model, and their difference scores were calculated. Both the threshold setting and final performance evaluation were conducted using this test data. In principle, outliers should exhibit larger difference scores compared to normal data. The binary decision threshold for identifying outliers was determined by selecting the threshold that achieved the best classification performance, as measured by the Area Under the Receiver Operating Characteristic (AUROC) [238] . Images with difference scores exceeding the decision threshold were classified as outliers. While some studies [239] ,[240] use cross-validation to fine-tune the threshold, collecting a sufficiently large validation set can be difficult, especially when anomalous samples are rare and challenging to gather. Thus, relying on the test data for threshold determination is often a practical alternative.

5.4 Results

The testing data consisted of 42 normal CBF images and 30 outliers. The distributions of three types of difference scores estimated by KDE, categorized by data type, are shown in Figure 5.4.1. Among the five artifacts, delayed arrival, poor signal, and poor labelling exhibited a significantly wide range of difference scores. In contrast, the distributions for motion and low contrast artifacts were more similar to

those of the normal images.

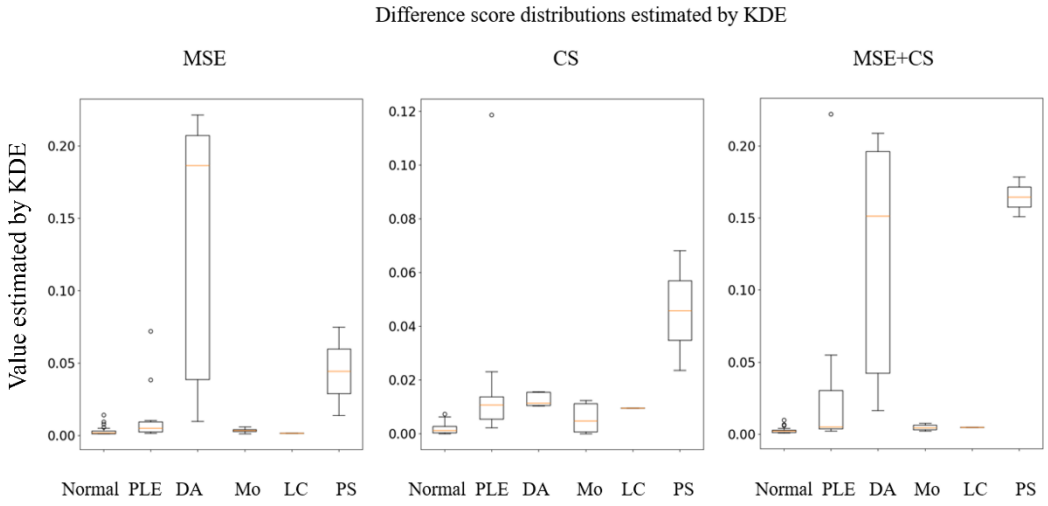


Figure 5.4.1 Box plots of different difference score distributions across the test dataset (42 normal and 30 outliers) estimated by KDE. MSE: mean squared error; CS: cosine similarity. PLE: poor labelling efficiency; DA: delayed arrival; Mo: motion; LC: low contrast; PS: poor signal.

The AUROC curves, which were used to evaluate classification performance based on three different types of difference scores, are presented in Figure 5.4.2. The AUROC scores for using cosine similarity and mean square error as the difference scores were 0.82 and 0.89, respectively. The highest AUROC score, 0.92, was achieved by combining mean square error and cosine similarity, indicating that this combination provided the best classification performance.

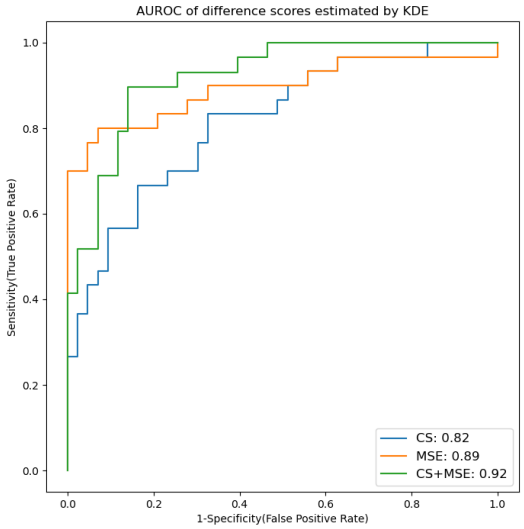


Figure 5.4.2 AUROC curves of difference scores used for testing. *MSE: mean squared error; CS: cosine similarity. Among them, using both CS+MSE achieves the highest AUROC of 0.92.*

Figure 5.4.3 displays the distribution of difference scores, computed using cosine similarity and mean squared error, as estimated by KDE. While some outliers exhibit distinctly high values and normal images consistently have low values, the lower-value region is densely populated with a mix of both normal and outlier images, which challenges the decision of the binary classification boundary.

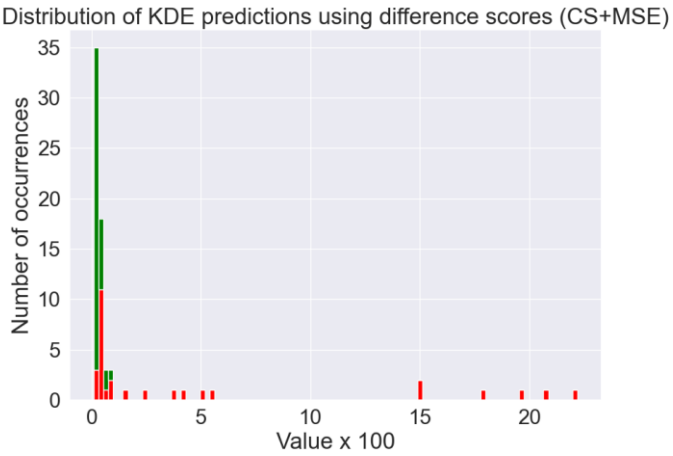


Figure 5.4.3 The distribution of difference scores from test data, which was predicted by KDE using cosine similarity and mean squared error. Red: outliers; Green: normal images.

The classification results based on the optimal decision threshold are depicted in Figure 5.4.4.

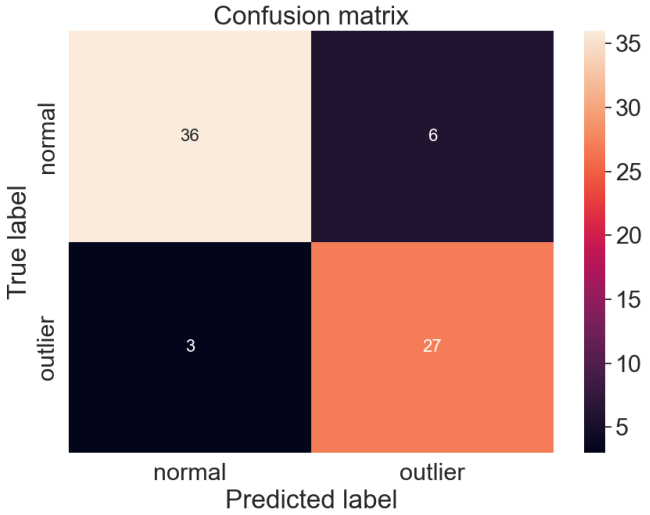


Figure 5.4.4 The classification results using the optimal threshold with the difference scores (mean square error and cosine similarity).

The three false negative (undetected) outliers shown in Figure 5.4.5 exhibit subtle artifacts. Additionally, six false positives (normal images misclassified as outliers) are shown in Figure 5.4.6. Although these images are labelled as normal, some artifacts are still present.

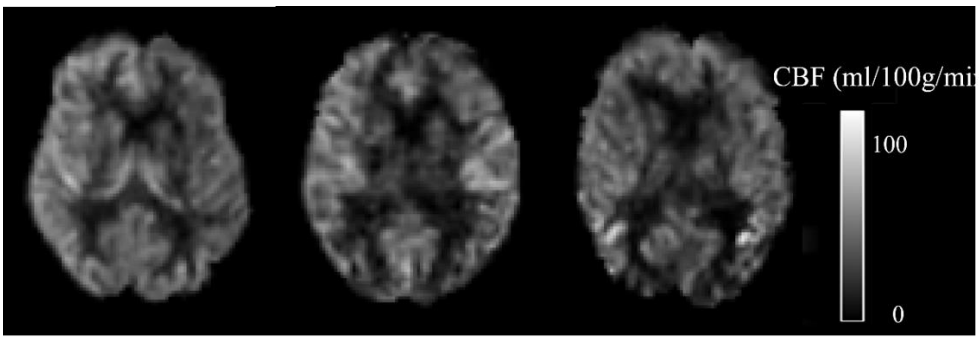


Figure 5.4.5 Three undetected outliers of CBF maps with slight artifacts (from left to right): poor labelling, motion, and motion.

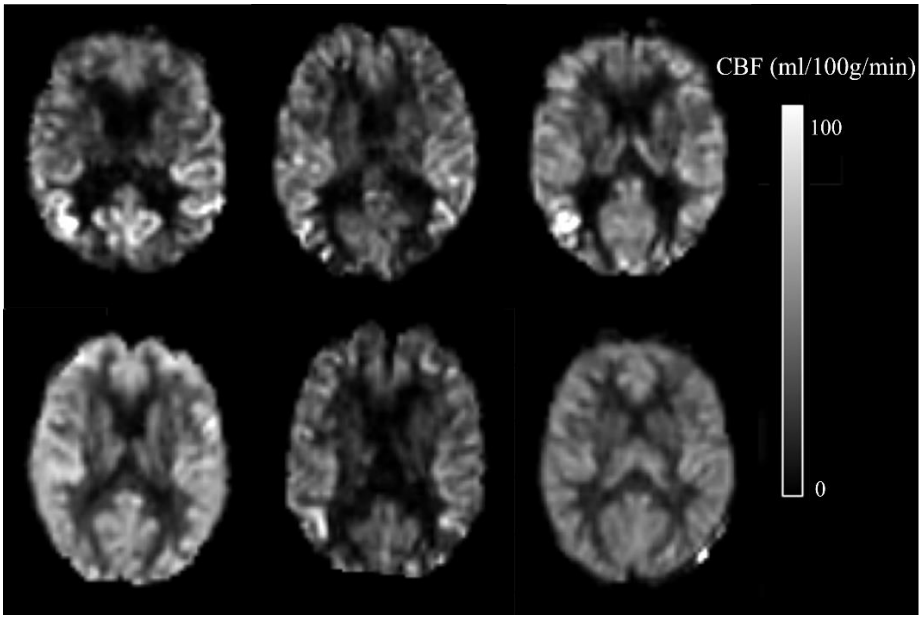


Figure 5.4.6 Six misclassified CBF images. Although these images were labelled as normal, some of them still contain artifacts. The top three images display delayed arrival of the ASL signal, while the bottom three CBF maps (from left to right) are normal, delayed arrival, and normal with a bit artifact at the right bottom rim due to imperfect brain extraction.

5.5 Discussion

In this chapter, QC was conducted for the ASL data using deep learning techniques. The VAE-GAN model was used to learn the pattern of the normal data, that use a feature-wise error replacing the element-wise error in conventional VAE model. The differences scores were computed between input data and reconstructed data including mean squared error, cosine similarity, and their combinations. Then KDE was used to measure the distribution of the difference scores, and thresholds were determined to identify outliers with high difference scores.

The combination of MSE and cosine similarity as difference scores yielded the highest performance, with an AUROC of 0.92. Using the optimal decision threshold based on this result, the model achieved the specificity of 0.9, successfully identifying most of the labelled outliers. Among the incorrectly classified normal images, although labelled as normal, artifacts were present in many of them, demonstrating the robustness and reduced sensitivity to manual labelling errors. Furthermore, artifacts such as the border zone sign and poor signal were found in the incorrectly classified normal images, highlighting the effectiveness of the model in capturing subtle features that were missed by the machine learning model in the previous chapter (see Figure 5.5.1), despite a lower sensitivity of the deep learning method. This also illustrates the model tolerance for labelling errors, showing the strength of the deviation-based approach, which does not rely on perfectly accurate labels. Additionally, even with a limited number of labelled outliers, the model was able to detect other types of artifacts, a task that would be nearly impossible for the machine learning model to accomplish.

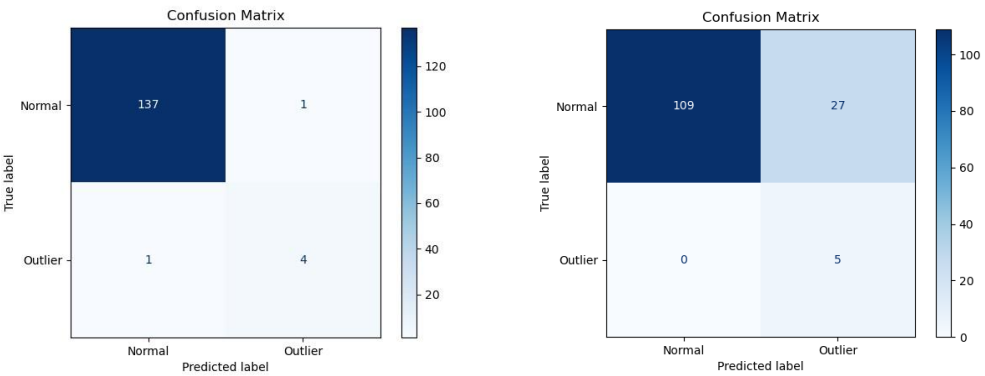


Figure 5.5.1 The comparison of the machine learning method in chapter 4 (left) and the deep learning method (right) on the same test data from chapter 4. Left: machine learning method (specificity=0.8, sensitivity=0.99); right: deep learning method (specificity=1, sensitivity=0.79).

There are several options for determining the decision thresholds for the difference scores. As observed in Figure 5.4.1, the distribution near the decision threshold used in this study is quite dense. Therefore, in practice, the criteria for setting the threshold should be carefully considered. In some cases, rather than aiming for the best overall model performance, the priority might be to avoid missing any outliers. An alternative approach would be to set the threshold based on a fixed percentage of the dataset, like in study [241] , such as filtering out the top 10% with the highest difference scores. This could ensure that low-quality data will not be overlooked, while only a small additional number of images will need to be manually examined.

Although the model successfully achieved the QC goal in ASL data, several issues remain worth discussing. One notable concern is the limited number of outliers in the dataset, which may affect the reliability of the model. The effectiveness of the model still needs to be verified with more outliers. However, collecting and simulating artifact-laden data poses a challenge, as images with evident errors are often discarded either during acquisition or pre-processing. Another aspect to consider is the number of normal images used for model training. In this study, all available normal images were utilized to construct the normal ASL patterns. However, for applying this model to other ASL datasets or adapting similar methodologies, the ideal number of training images should be investigated. Insufficient normal data may prevent the deep learning model from adequately learning normal patterns, while using too many images could introduce the risk of including lower-quality images, which might bias the training process. Although the model demonstrated robustness to label inconsistencies, this issue remains a potential area for further refinement and improvement.

There are several limitations in this study. First, the CBF maps in the standard space were used for consistency in model training, which, while facilitating image preparation, may have introduced geometrical distortions. This could have caused discrepancies between the CBF maps in the native and the standard space, potentially affecting the QC results. Additionally, given the 8 mm slice thickness and the 1 mm inter-slice gap of the CBF maps, using 2D slices from sagittal or coronal planes was avoided. Moreover, by using only the central axial slices, the risk of overfitting was minimized, as the focus was placed on key anatomical structures such as CSF, white matter, and primary cortical areas, which are essential for assessing CBF map quality.

While the results demonstrated that using the middle axial slice alone is sufficient to identify most artifacts in the current dataset, this approach could overlook perfusion information from the superior or inferior slices, potentially limiting the model generalizability to other datasets. Regarding the resolution of resampling, the downsampling chosen in this study was closer to the original size of the axial slices (80x80), which balances anatomical relevance with computational efficiency. While upsampling (e.g., scaling to 128x128) could provide more detail, it introduces interpolation that conflicts with the inherently low-resolution nature of ASL data. The choice of a lower resolution strikes a reasonable balance.

Further research is needed to validate these choices and explore the use of 3D models to better represent the inherent 3D structure of CBF maps. This would provide a more comprehensive assessment of the data, and metrics like Dice scores could be employed as additional difference scores. Combining machine learning with deep learning approaches could further improve QC, as incorporating specific ASL QC metrics as features may help in more accurately identifying outliers and enhancing decision-making robustness.

5.6 Conclusion

In this chapter, deep learning techniques were employed to develop a robust model conducting automatic QC on ASL CBF maps, filling the gap of automatic QC for CBF maps using ASL MRI on a relatively large cohort. The model requires only a limited number of labelled images and does not depend on highly precise image labelling, yet it achieved higher accuracy and specificity compared to traditional machine learning methods used in the chapter 4. In practical applications, the decision threshold for identifying outliers can be adjusted to capture more potential low-quality images, providing flexibility in balancing sensitivity and specificity in QC tasks.

6 Conclusion and Future Directions

6.1 Conclusion

This thesis has addressed two critical challenges in ASL MRI: the application of partial volume effects correction for CBF measurements in a large ASL dataset and the development of standardized QC protocols for ASL data. Additionally, automated QC tools were developed, with the aim of improving the accuracy and efficiency of outlier detection, ensuring more reliable ASL data for clinical and research applications.

Chapter 3 of this thesis investigated regional CBF changes with age and examined how correcting for PVEs, which accounts for changes in grey matter, might affect the understanding of CBF decline in a large aging population. In addition to the conventional volumetric pipeline, a surface-based ASL pre-processing pipeline was developed to account for the intrinsic structure of cortex. These pipelines can serve as templates for future ASL data processing. The findings revealed that while global decreases in grey matter CBF are observed with aging before PVEc, the corrected data showed more regionally selective perfusion changes. This suggests that that variations in cerebral perfusion with age observed with imaging are influenced by regional changes in anatomy that can be accommodated with PVEc, but perfusion variations are still observable even after PVE is accounted for.

Chapter 4 focused on developing standardized QC protocols specific to ASL MRI. These protocols incorporated metrics related to signal quality, perfusion quality, and registration quality. Additionally, interactive QC reports were developed to facilitate manual screening. The effectiveness of individual QC metrics was evaluated using machine learning, and the proposed metrics were combined to train a machine model to identify outliers, which achieved high performance despite the limited number of outliers in the test data.

In Chapter 5, deep learning techniques were employed to develop an automated QC tool for ASL data. A VAE-GAN model was trained to recognize patterns in normal ASL data and detect outliers by measuring the differences between ASL images and the learned normal pattern. These differences were modelled by Kernel Density Estimation, and images with difference scores below a pre-defined decision threshold

were classified as outliers. This method outperformed traditional machine learning models in identifying subtle anomalies. Additionally, the VAE-GAN model demonstrated robustness and effectiveness by reducing the reliance on precise manual labels and accommodating variability in ASL data. This highlights its potential as an efficient and reliable automated QC tool for ASL imaging.

In conclusion, the findings of this thesis contribute the foundation for future advancements in both the application of partial volume effects correction and the standardized QC protocols in ASL MRI. The automated QC tools developed, either using ASL quality metrics with machine learning or deep learning techniques, have demonstrated strong potential in identifying outliers in ASL data. These tools hold promise for significantly enhancing the clinical utility of ASL MRI, paving the way for more reliable and effective imaging in both research and clinical settings.

6.2 Future Directions

While each chapter has discussed the limitations and potential solutions to improve upon the research presented in this thesis, several key points emerge as future directions based on the findings.

In Chapter 3, partial volume effects were corrected to account for the potential anatomical changes in GM volumes with aging. For non-PVEc ASL data, an 80% GM partial volume threshold was used during pre-processing to define "pure" grey matter for calculating GM CBF. This high threshold was intended to minimize the sensitivity of non-PVEc GM CBF to PVE, however, this approach may have contributed to the subtle or non-evident age-related CBF decline observed when compared to PVEc results. Moving forward, a lower GM partial volume threshold, such as 50%, could be used to capture more GM tissue that might be significantly influenced by PVE.

In Chapter 4, quality metrics specific to ASL MRI were investigated and evaluated using machine learning. These metrics were primarily derived from raw ASL images (label-control image pairs) and CBF maps in different spaces. However, there are commonly ancillary images associated with ASL data, such as arterial transit time maps. Although ATT maps were not used in this study, given the single PLD data,

future QC metrics should consider these ancillary images to account for various ASL sequences. Additionally, general image quality metrics and other registration quality metrics might also prove useful for ASL QC. However, given the potentially large number of general quality metrics, further investigation and assessment are warranted to determine which are most relevant for ASL QC.

While the automated QC tools developed in this thesis demonstrated strong performance using a large ASL dataset, it is crucial to validate their effectiveness on ASL images from different sources. To support broader adoption in research, developing and publishing user-friendly Python packages that offer flexible input from various ASL pre-processing tools would be beneficial. Despite the success of automated tools, visual inspection remains essential, underscoring the value of QC reports. In this thesis, QC reports were developed using Jupyter Notebook to facilitate manual checks of ASL images. Although the reports support interactive controls, more user-friendly formats, such as HTML, would be advantageous as they are easier to share. Enhancing the usability and aesthetics of these QC reports is also an important future direction.

In Chapter 5, it was discussed that further improvements can be made to the VAE-GAN model. Beyond the deep learning approach presented, prior studies [107] [108] proposed alternative methods such as structural correlation-based outlier rejection algorithms and empirical robust Bayesian methods for ASL QC. These methods leverage information from structural images to identify outliers more effectively. Building on this idea, incorporating structural images into the deep learning model could enhance its capabilities by providing additional context beyond CBF maps alone. In future work, the performance of partial volume segmentations could be integrated into the model as a sub-task of the neural networks, contributing to a more comprehensive quality assessment of ASL data.

Moreover, combining machine learning with deep learning techniques may offer further improvements. For instance, the QC metrics developed in Chapter 4 could be incorporated as features in the deep learning model, enhancing its ability to distinguish outliers from normal data. This hybrid approach could leverage the strengths of both methodologies, potentially offering a more robust and precise ASL QC system.

For the clinical application of ASL QC, an ultimate goal is to integrate QC processes directly into the scanners, enabling real-time evaluation of image quality immediately after acquisition. This integration would allow clinicians to assess the reliability of the acquired data on the spot, facilitating immediate decision-making on whether a scan needs to be repeated or adjusted, ensuring optimal data quality for accurate diagnosis or treatment planning. Based on the research of this thesis, one potential solution is to embed machine learning algorithms alongside ASL pre-processing within the scanner software. This approach would enable quality metrics to be calculated from processed ASL data and then fed into machine learning algorithms for real-time QC. However, a significant challenge is that ASL pre-processing, particularly for complex kinetic models, can be computation-intensive and time-consuming, making real-time application difficult. This challenge, however, is not unsolvable. For example, a prior study by Y. Zhang [243] proposed using neural networks to replace the parameter estimation step in kinetic models for CBF measurements in ASL data. This method significantly accelerated the process of CBF quantification while maintaining comparable accuracy to conventional approaches like Variational Bayes. By reducing the computational cost, such advancements make it feasible to integrate machine learning-based QC processes directly into scanners. While the implementation of these solutions would be technically demanding, with hurdles such as hardware limitations, data processing speed, and the development of advanced algorithms, their realization would mark a significant advancement in the clinical utility of ASL MRI. Overcoming these challenges could enable real-time ASL QC, improve clinical workflow efficiency, and ensure consistently high-quality data for diagnosis and treatment planning.

Bibliography

[1] Alsop, D. C., Detre, J. A., Golay, X., Günther, M., Hendrikse, J., Hernandez-Garcia, L., Lu, H., MacIntosh, B. J., Parkes, L. M., Smits, M., van Osch, M. J., Wang, D. J., Wong, E. C., & Zaharchuk, G. (2015). 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. Magnetic resonance in medicine, 73(1), 102–116. https://doi.org/10.1002/mrm.25197

[2] Akgoren N, Fabricius M, Lauritzen M. 1994. Importance of nitric oxide for local increases of blood flow in rat cerebellar cortex during electrical stimulation. Proc Natl Acad Sci U S A 91:5903–5907.

[3] Hoge RD, Pike GB. 2001. Oxidative metabolism and the detection of neuronal activation via imaging. J Chem Neuroanat 22:43–52.

[4] Chappell, M. A., McConnell, F. A. K., Golay, X., Günther, M., Hernandez-Tamames, J. A., van Osch, M. J., & Asllani, I. (2021). Partial volume correction in arterial spin labeling perfusion MRI: A method to disentangle anatomy from physiology or an analysis step too far?. NeuroImage, 238, 118236. https://doi.org/10.1016/j.neuroimage.2021.118236

[5] Amukotuwa SA, Yu C, Zaharchuk G. 3D Pseudocontinuous arterial spin labeling in routine clinical practice: A review of clinically significant artifacts. J Magn Reson Imaging. 2016 Jan;43(1):11-27. doi: 10.1002/jmri.24873. Epub 2015 Apr 9. PMID: 25857715.

[6] Mutsaerts, H. J., van Dalen, J. W., Heijtel, D. F., Groot, P. F., Majoie, C. B., Petersen, E. T., Richard, E., & Nederveen, A. J. (2015). Cerebral Perfusion Measurements in Elderly with Hypertension Using Arterial Spin Labeling. PloS one, 10(8), e0133717. https://doi.org/10.1371/journal.pone.0133717

[7] Asllani, I., Borogovac, A., Brown, T.R., 2008. Regression algorithm correcting for partial volume effects in arterial spin labeling MRI. Magn. Reson. Med. 60, 1362–1371. https://doi.org/10.1002/mrm.21670.

[8] Chappell, M. A., Groves, A. R., MacIntosh, B. J., Donahue, M. J., Jezzard, P., & Woolrich, M. W. (2011). Partial volume correction of multiple inversion time arterial spin labeling MRI data. Magnetic resonance in medicine, 65(4), 1173–1183. https://doi.org/10.1002/mrm.22641

[9] Alfaro-AlmagroF, Jenkinson M, BangerterN, AnderssonJ,Griffanti L, DouaudG, et al. UK Biobank Brain Imaging: Automated Processing Pipeline and Quality Control for 100,000 subjects. In: Organiza tion for Human Brain Mapping. Geneve,Switzerland; 2016. p. 1877.

[10] F. Bloch, W. Hansen, and E. Martin, Packard m. nuclear induction, Physical Review 127 (1946)

[11] Khashami, F. (2024). A mini review of NMR and MRI. *arXiv preprint arXiv:2401.01389*.

[12] A. Abragam, The principles of nuclear magnetism, 32 (Oxford university press, 1983).

[13] S. C. Bushong and G. Clarke, Magnetic Resonance Imaging-E-Book: Physical and Biological Principles (Elsevier Health Sciences, 2013).

[14] Haskell MW, Nielsen JF, Noll DC. Off-resonance artifact correction for MRI: A review. NMR Biomed. 2023 May;36(5):e4867. doi: 10.1002/nbm.4867. Epub 2022 Dec 14. PMID: 36326709; PMCID: PMC10284460.

[15] Hahn EL. Spin Echoes. Physical Review. 1950;80(4):580-594.

[16] Elster AD. Gradient-echo MR imaging: techniques and acronyms. Radiology. 1993;186(1):1-8.

[17] Stanisz GJ, Odrobina EE, Pun J, et al. T1, T2 relaxation and magnetization transfer in tissue at 3T. Magn Reson Med. 2005;54(3):507-512.

[18] Gallagher TA, Nemeth AJ, Hacein-Bey L. An introduction to the Fourier transform: relationship to MRI. AJR Am J Roentgenol. 2008 May;190(5):1396-405. doi: 10.2214/AJR.07.2874. PMID: 18430861.

[19] Frisoni, G. B., Fox, N. C., Jack Jr, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature reviews neurology*, *6*(2), 67-77.

[20] Berger A. (2002). Magnetic resonance imaging. BMJ (Clinical research ed.), 324(7328), 35. https://doi.org/10.1136/bmj.324.7328.35

[21] Pohost, G. M., Elgavish, G. A., & Evanochko, W. T. (1986). Nuclear magnetic resonance imaging: with or without nuclear?. *Journal of the American College of Cardiology*, *7*(3), 709–710. https://doi.org/10.1016/s0735-1097(86)80486-7

[22] Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL; Brain Development Cooperative Group. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage. 2011 Jan 1;54(1):313-27. doi: 10.1016/j.neuroimage.2010.07.033. Epub 2010 Jul 23. PMID: 20656036; PMCID: PMC2962759.

[23] MEIER, P., & ZIERLER, K. L. (1954). On the theory of the indicator-dilution method for measurement of blood flow and volume. *Journal of applied physiology*, *6*(12), 731–744. https://doi.org/10.1152/jappl.1954.6.12.731

[24] Chappell MA, MacIntosh BJ, Donahue MJ, Günther M, Jezzard P, Woolrich MW. Separation of macrovascular signal in multi-inversion time arterial spin labelling MRI. Magn Reson Med. 2010 May;63(5):1357-65. doi: 10.1002/mrm.22320. PMID: 20432306.

[25] Parkes LM. Quantification of cerebral perfusion using arterial spin labeling: two-compartment models. Journal of magnetic resonance imaging : JMRI. 2005; 22(6):732–736. [PubMed: 16267854]

[26] Pinto J, Chappell MA, Okell TW, Mezue M, Segerdahl AR, Tracey I, Vilela P, Figueiredo P. Calibration of arterial spin labeling data-potential pitfalls in post-processing. Magn Reson Med. 2020 Apr;83(4):1222-1234. doi: 10.1002/mrm.28000. Epub 2019 Oct 12. PMID: 31605558; PMCID: PMC6972489.

[27] Hernandez-Garcia L, Aramendía-Vidaurreta V, Bolar DS, et al. Recent Technical Developments in ASL: A Review of the State of the Art. *Magn Reson Med*. 2022; 88: 2021-2042. doi:10.1002/mrm.29381

[28] Paschal CB, Morris HD. K-space in the clinic. J Magn Reson Imaging. 2004 Feb;19(2):145-59. doi: 10.1002/jmri.10451. PMID: 14745747.

[29] Dhabalia R, Kashikar SV, Parihar PS, Mishra GV. Unveiling the Intricacies: A Comprehensive Review of Magnetic Resonance Imaging (MRI) Assessment of T2-Weighted Hyperintensities in the Neuroimaging Landscape. Cureus. 2024 Feb 24;16(2):e54808. doi: 10.7759/cureus.54808. PMID: 38529430; PMCID: PMC10961652.

[30] Mezrich R. A perspective on K-space. Radiology. 1995;195(2):297-315.

[31] Zhang, L. X. Improving the Quantification of Cerebral Haemodynamics Using Optimally Sampled Arterial Spin Labelling MRI. University of Oxford, 2022.

[32] Huang SY, Seethamraju RT, Patel P, Hahn PF, Kirsch JE, Guimaraes AR. Body MR Imaging: Artifacts, k-Space, and Solutions. Radiographics. 2015 Sep-Oct;35(5):1439-60. doi: 10.1148/rg.2015140289. Epub 2015 Jul 24. Erratum in: Radiographics. 2015 Sep-Oct;35(5):1624. PMID: 26207581; PMCID: PMC4613875.

[33] Mark Jenkinson and Michael Chappell, “Introduction to Neuroimaging Analysis “, (Oxford: Oxford University Press), 2018, 276 pages, ISBN: 978-0198816300.

[34] Chappell, M., MacIntosh, B., Okell, T., 2018. Introduction to Perfusion Quantification Using Arterial Spin Labelling. Oxford University Press.

[35] Jbabdi, S., & Johansen-Berg, H. (2011). Tractography: where do we go from here?. Brain connectivity, 1(3), 169–183. https://doi.org/10.1089/brain.2011.0033

[36] DeLaPaz R. L. (1994). Echo-planar imaging. Radiographics : a review publication of the Radiological Society of North America, Inc, 14(5), 1045–1058. https://doi.org/10.1148/radiographics.14.5.7991813

[37] Glover G. H. (2012). Spiral imaging in fMRI. NeuroImage, 62(2), 706–712. https://doi.org/10.1016/j.neuroimage.2011.10.039

[38] Glover G. H. (2011). Overview of functional magnetic resonance imaging. Neurosurgery clinics of North America, 22(2), 133–vii. https://doi.org/10.1016/j.nec.2010.11.001

[39] Jahng, G. H., Li, K. L., Ostergaard, L., & Calamante, F. (2014). Perfusion magnetic resonance imaging: A comprehensive update on principles and techniques. Korean Journal of Radiology, 15(5), 554-577. https://doi.org/10.3348/kjr.2014.15.5.554

[40] Merkle, F., Haupt, B., El-Essawi, A., & Hetzer, R. (2012). State of the art in cardiovascular perfusion: now and in the next decade. HSR proceedings in intensive care & cardiovascular anesthesia, 4(4), 211–216.

[41] Pichler, B. J., Wehrl, H. F., Kolb, A., & Judenhofer, M. S. (2008). Positron emission tomography/magnetic resonance imaging: the next generation of multimodality imaging?. Seminars in nuclear medicine, 38(3), 199–208. https://doi.org/10.1053/j.semnuclmed.2008.02.001

[42] Boxerman, J. L., Quarles, C. C., Hu, L. S., Erickson, B. J., Gerstner, E. R., Smits, M., Kaufmann, T. J., Barboriak, D. P., Huang, R. H., Wick, W., Weller, M., Galanis, E., Kalpathy-Cramer, J., Shankar, L., Jacobs, P., Chung, C., van den Bent, M. J., Chang, S., Al Yung, W. K., Cloughesy, T. F., … Jumpstarting Brain Tumor Drug Development Coalition Imaging Standardization Steering Committee (2020). Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in high-grade gliomas. Neuro-oncology, 22(9), 1262–1275. https://doi.org/10.1093/neuonc/noaa141

[43] Gordon, Y., Partovi, S., Müller-Eschner, M., Amarteifio, E., Bäuerle, T., Weber, M. A., Kauczor, H. U., & Rengier, F. (2014). Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. Cardiovascular diagnosis and therapy, 4(2), 147–164. https://doi.org/10.3978/j.issn.2223-3652.2014.03.01

[44] Liu W, Lou X, Ma L. (2016). Use of 3D pseudo-continuous arterial spin labeling to characterize sex and age differences in cerebral blood flow. Neuroradiology. 58:943–48. https://doi.org/10.1007/s00234-016-1713-y PMID:27380039

[45] Preibisch, C., Sorg, C., Förschler, A., Grimmer, T., Sax, I., Wohlschläger, A. M., Perneczky, R., Förstl, H., Kurz, A., Zimmer, C., & Alexopoulos, P. (2011). Age-related cerebral perfusion changes in the parietal and temporal lobes measured by pulsed arterial spin labeling. Journal of magnetic resonance imaging : JMRI, 34(6), 1295–1302. https://doi.org/10.1002/jmri.22788

[46] Lee, C., Lopez, O. L., Becker, J. T., Raji, C., Dai, W., Kuller, L. H., & Gach, H. M. (2009). Imaging cerebral blood flow in the cognitively normal aging brain with arterial spin labeling: implications for imaging of neurodegenerative disease. Journal of neuroimaging : official journal of the American Society of Neuroimaging, 19(4), 344–352. https://doi.org/10.1111/j.1552-6569.2008.00277.x

[47] Lu H, Clingman C, Golay X, van Zijl PC. Determining the longitudinal relaxation time (T1) of blood at 3.0 Tesla. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine. 2004; 52(3):679–682.

[48] Zhang X, Petersen ET, Ghariq E, De Vis JB, Webb AG, Teeuwisse WM, Hendrikse J, van Osch MJ. In vivo blood T(1) measurements at 1.5 T, 3 T, and 7 T. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine. 2012 10. 1002/mrm.24550.

[49] Petersen ET, Mouridsen K, Golay X. The QUASAR reproducibility study, Part II: Results from a multi-center Arterial Spin Labeling test-retest study. NeuroImage. 2010; 49(1):104–113. [PubMed: 19660557]

[50] Bokkers RP, van der Worp HB, Mali WP, Hendrikse J. Noninvasive MR imaging of cerebral perfusion in patients with a carotid artery stenosis. Neurology. 2009; 73(11):869–875. [PubMed: 19752454]

[51] Qiu M, Paul Maguire R, Arora J, Planeta-Wilson B, Weinzimmer D, Wang J, Wang Y, Kim H, Rajeevan N, Huang Y, Carson RE, Constable RT. Arterial transit time effects in pulsed arterial spin labeling CBF mapping: insight from a PET and MR study in normal human subjects. Magn Reson Med. 2010 Feb;63(2):374-84. doi: 10.1002/mrm.22218. PMID: 19953506; PMCID: PMC2867043.

[52] Williams DS, Detre JA, Leigh JS, Koretsky AP. Magnetic resonance imaging of perfusion using spin inversion of arterial water. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89(1):212–216. [PubMed: 1729691]

[53] Kim SG. Quantification of relative cerebral blood flow change by flow-sensitive alternating inversion recovery (FAIR) technique: application to functional mapping. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine. 1995; 34(3):293–301.

[54] Petcharunpaisan, S., Ramalho, J., & Castillo, M. (2010). Arterial spin labeling in neuroimaging. World journal of radiology, 2(10), 384–398. https://doi.org/10.4329/wjr.v2.i10.384

[55] Trampel R, Jochimsen TH, Mildner T, Norris DG, Möller HE. Efficiency of flow-driven adiabatic spin inversion under realistic experimental conditions: a computer simulation. Magn Reson Med. 2004 Jun;51(6):1187-93. doi: 10.1002/mrm.20080. PMID: 15170839.

[56] Wong EC, Buxton RB, Frank LR. Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II).Magn Reson Med 1998;39:702-708.

[57] Noguchi T, Nishihara M, Hara Y, Hirai T, Egashira Y, Azama S, Irie H. A technical perspective for understanding quantitative arterial spin-labeling MR imaging using Q2TIPS. Magn Reson Med Sci. 2015;14(1):1-12. doi: 10.2463/mrms.2013-0064. Epub 2014 Dec 15. PMID: 25500774.

[58] Borogovac A, Asllani I. Arterial Spin Labeling (ASL) fMRI: advantages, theoretical constrains, and experimental challenges in neurosciences. Int J Biomed Imaging. 2012;2012:818456. doi:

10.1155/2012/818456. Epub 2012 Feb 22. Erratum in: Int J Biomed Imaging. 2012;2012:658101. PMID: 22966219; PMCID: PMC3432878.

[59] Maleki N, Dai W, Alsop DC. Optimization of background suppression for arterial spin labeling perfusion imaging. MAGMA. 2012; 25(2):127–133. [PubMed: 22009131]

[60] Garcia DM, Duhamel G, Alsop DC. Efficiency of inversion pulses for background suppressed arterial spin labeling. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine. 2005; 54(2):366–372.

[61] Vidorreta M, Wang Z, Rodriguez I, Pastor MA, Detre JA, Fernandez-Seara MA. Comparison of 2D and 3D single-shot ASL perfusion fMRI sequences. NeuroImage. 2012; 66C:662–671. [PubMed: 23142069]

[62] Nielsen JF, Hernandez-Garcia L. Functional perfusion imaging using pseudocontinuous arterial spin labeling with low-flip-angle segmented 3D spiral readouts. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine. 2013; 69(2):382–390.

[63] Chang, Y. V., Vidorreta, M., Wang, Z., & Detre, J. A. (2017). 3D-accelerated, stack-of-spirals acquisitions and reconstruction of arterial spin labeling MRI. Magnetic resonance in medicine, 78(4), 1405–1419. https://doi.org/10.1002/mrm.26549

[64] Nachmani A, Schurr R, Joskowicz L, Mezer AA. The effect of motion correction interpolation on quantitative T1 mapping with MRI. Med Image Anal. 2019 Feb;52:119-127. doi: 10.1016/j.media.2018.11.012. Epub 2018 Dec 1. PMID: 30529225.

[65] Tsao J. (2010). Ultrafast imaging: principles, pitfalls, solutions, and applications. Journal of magnetic resonance imaging : JMRI, 32(2), 252–266. https://doi.org/10.1002/jmri.22239

[66] Grade, M., Hernandez Tamames, J. A., Pizzini, F. B., Achten, E., Golay, X., & Smits, M. (2015). A neuroradiologist's guide to arterial spin labeling MRI in clinical practice. Neuroradiology, 57(12), 1181–1202. https://doi.org/10.1007/s00234-015-1571-z

[67] Madai, V. I., Martin, S. Z., von Samson-Himmelstjerna, F. C., Herzig, C. X., Mutke, M. A., Wood, C. N., Thamm, T., Zweynert, S., Bauer, M., Hetzer, S., Günther, M., & Sobesky, J. (2016). Correction for Susceptibility Distortions Increases the Performance of Arterial Spin Labeling in Patients with Cerebrovascular Disease. Journal of neuroimaging : official journal of the American Society of Neuroimaging, 26(4), 436–444. https://doi.org/10.1111/jon.12331

[68] Rong, Y., Rosu-Bubulac, M., Benedict, S. H., Cui, Y., Ruo, R., Connell, T., Kashani, R., Latifi, K., Chen, Q., Geng, H., Sohn, J., & Xiao, Y. (2021). Rigid and Deformable Image Registration for Radiation Therapy: A Self-Study Evaluation Guide for NRG Oncology Clinical Trial

Participation. Practical radiation oncology, 11(4), 282–298. https://doi.org/10.1016/j.prro.2021.02.007

[69] Crum, W. R., Hartkens, T., & Hill, D. L. (2004). Non-rigid image registration: theory and practice. The British journal of radiology, 77 Spec No 2, S140–S153. https://doi.org/10.1259/bjr/25329214

[70] Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *neuroimage*, *46*(1), 39-46.

[71] Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825-841.

[72] Buxton, R. B., Frank, L. R., Wong, E. C., Siewert, B., Warach, S., & Edelman, R. R. (1998). A general kinetic model for quantitative perfusion imaging with arterial spin labeling. Magnetic resonance in medicine, 40(3), 383–396. https://doi.org/10.1002/mrm.1910400308

[73] Mutsaerts, H.J.M.M., Richard, E., Heijtel, D.F.R., Van Osch, M.J.P., Majoie, C.B.L.M., Nederveen, A.J., 2014. Gray matter contamination in arterial spin labeling white matter perfusion measurements in patients with dementia. NeuroImage Clin. 4, 139–144. https://doi.org/10.1016/j.nicl.2013.11.003. Review Article.

[74] Golay, X., Hendrikse, J., Lim, T.C.C., 2004. Perfusion imaging using arterial spin labeling. Top. Magn. Reson. Imaging 15, 10–27. https://doi.org/10.1097/00002142 200402000-00003.

[75] Johnson, N.A., Jahng, G.-H., Weiner, M.W., Miller, B.L., Chui, H.C., Jagust, W.J., Gorno Tempini, M.L., Schuff, N., 2005. Pattern of cerebral hypoperfusion in Alzheimer disease and mild cognitive impairment measured with arterial spin-labeling MR imaging: initial Experience1. Radiology 234, 851–859. https://doi.org/10.1148/ radiol.2343040197.

[76] Le Heron, C.J., Wright, S.L., Melzer, T.R., Myall, D.J., Macaskill, M.R., Livingston, L., Keenan, R.J., Watts, R., Dalrymple-Alford, J.C., Anderson, T.J., 2014. Comparing cerebral perfusion in Alzheimer's disease and Parkinson's disease dementia: an ASL MRI study. J. Cereb. Blood Flow. Metab. 34, 1–7. https://doi.org/10.1038/ jcbfm.2014.40.

[77] Henery, C.C., Mayhew, T.M., 1989. The cerebrum and cerebellum of the fixed human brain: efficient and unbiased estimates of volumes and cortical surface areas. J. Anat. 167, 167–180.

[78] Steketee, R.M.E., Bron, E.E., Meijboom, R., Houston, G.C., Klein, S., Mutsaerts, H.J.M.M., Mendez Orellana, C.P., de Jong, F.J., van Swieten, J.C., van der Lugt, A., Smits, M., 2015. Early-stage differentiation between presenile Alzheimer's disease and frontotemporal dementia using arterial spin labeling MRI. Eur. Radiol. https:// doi.org/10.1007/s00330-015-3789-x.

[79] Zhao, M.Y., Petersen, E.T., Chappell, M.A., 2016. Cerebral perfusion quantification using Turbo-QUASAR arterial spin labelling MRI: a model-based approach. Proc. MEIBioeng 2–4.

[80] Groves, A.R., Chappell, M.A., Woolrich, M.W., 2009. Combined spatial and non-spatial prior for inference on MRI time-series. Neuroimage 45, 795–809. https://doi.org/ 10.1016/j.neuroimage.2008.12.027.

[81] Penny, W.D., Trujillo-Barreto, N.J., Friston, K.J., 2005. Bayesian fMRI time series analysis with spatial priors. Neuroimage 24, 350–362. https://doi.org/10.1016/ j.neuroimage.2004.08.034.

[82] Jaganmohan, D., Pan, S., Kesavadas, C., & Thomas, B. (2021). A pictorial review of brain arterial spin labelling artefacts and their potential remedies in clinical studies. The neuroradiology journal, 34(3), 154–168. https://doi.org/10.1177/1971400920977031

[83] Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PloS one, 12(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

[84] Cox R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and biomedical research, an international journal, 29(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

[85] Mutsaerts, H. J. M. M., Petr, J., Groot, P., Vandemaele, P., Ingala, S., Robertson, A. D., Václavů, L., Groote, I., Kuijf, H., Zelaya, F., O'Daly, O., Hilal, S., Wink, A. M., Kant, I., Caan, M. W. A., Morgan, C., de Bresser, J., Lysvik, E., Schrantee, A., Bjørnebekk, A., … Barkhof, F. (2020). ExploreASL: An image processing pipeline for multi-center ASL perfusion MRI studies. NeuroImage, 219, 117031. https://doi.org/10.1016/j.neuroimage.2020.117031

[86] Adebimpe, A., Bertolero, M., Dolui, S., Cieslak, M., Murtha, K., Baller, E. B., Boeve, B., Boxer, A., Butler, E. R., Cook, P., Colcombe, S., Covitz, S., Davatzikos, C., Davila, D. G., Elliott, M. A., Flounders, M. W., Franco, A. R., Gur, R. E., Gur, R. C., Jaber, B., … Satterthwaite, T. D. (2022). ASLPrep: a platform for processing of arterial spin labeled MRI and quantification of regional brain perfusion. Nature methods, 19(6), 683–686. https://doi.org/10.1038/s41592-022-01458-7

[87] Woodard, J. P., & Carley-Spencer, M. P. (2006). No-reference image quality metrics for structural MRI. Neuroinformatics, 4(3), 243–262. https://doi.org/10.1385/NI:4:3:243

[88] Mortamet, B., Bernstein, M. A., Jack, C. R., Jr, Gunter, J. L., Ward, C., Britson, P. J., Meuli, R., Thiran, J. P., Krueger, G., & Alzheimer's Disease Neuroimaging Initiative (2009). Automatic quality assessment in structural brain magnetic resonance imaging. Magnetic resonance in medicine, 62(2), 365–372. https://doi.org/10.1002/mrm.21992

[89] Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jr, Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology, 74(3), 201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25

[90] Pizarro, R. A., Cheng, X., Barnett, A., Lemaitre, H., Verchinski, B. A., Goldman, A. L., Xiao, E., Luo, Q., Berman, K. F., Callicott, J. H., Weinberger, D. R., & Mattay, V. S. (2016). Automated Quality Assessment of Structural Magnetic Resonance Brain Images Based on a Supervised Machine Learning Algorithm. Frontiers in neuroinformatics, 10, 52. https://doi.org/10.3389/fninf.2016.00052

[91] Kuncheva, L.I., 2004. Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons.

[92] Kittler, J., Hatef, M., Duin, R.P., Matas, J., 1998. On combining classifiers. IEEE Trans. Pattern Analysis Mach. Intell. 20 (3), 226–239.

[93] Lepping, R. J., Yeh, H. W., McPherson, B. C., Brucks, M. G., Sabati, M., Karcher, R. T., Brooks, W. M., Habiger, J. D., Papa, V. B., & Martin, L. E. (2023). Quality control in resting-state fMRI: the benefits of visual inspection. Frontiers in neuroscience, 17, 1076824. https://doi.org/10.3389/fnins.2023.1076824

[94] Lu, W., Dong, K., Cui, D., Jiao, Q., & Qiu, J. (2019). Quality assurance of human functional magnetic resonance imaging: a literature review. Quantitative imaging in medicine and surgery, 9(6), 1147–1162. https://doi.org/10.21037/qims.2019.04.18

[95] Reynolds, R. C., Taylor, P. A., & Glen, D. R. (2023). Quality control practices in FMRI analysis: Philosophy, methods and examples using AFNI. Frontiers in neuroscience, 16, 1073800. https://doi.org/10.3389/fnins.2022.1073800

[96] Shehzad Z, Giavasis S, Li Q, Benhajali Y, Yan C, Yang Z, Milham M, Bellec P and Craddock C (2015). The Preprocessed Connectomes Project Quality Assessment Protocol - a resource for measuring the quality of MRI data.. Front. Neurosci. Conference Abstract: Neuroinformatics 2015. doi: 10.3389/conf.fnins.2015.91.00047

[97] Atkinson D, Hill DL, Stoyle PN, Summers PE, Keevil SF (1997). Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. IEEE Trans Med Imaging. 16(6):903-10.

[98] Friedman, L., Stern, H., Brown, G. G., Mathalon, D. H., Turner, J., Glover, G. H., … & Potkin, S. G. (2008). Test–retest and between-site reliability in a multicenter fMRI study. Human brain mapping, 29(8), 958-972

[99] Nichols, T. (2012, Oct 28). Standardizing DVARS. Retrieved from http://blogs.warwick.ac.uk/nichols/entry/standardizing_dvars.

[100] Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142-2154.

[101] Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D. A., Gallagher, L., Kennedy, D. P., Keown, C. L., Keysers, C., Lainhart, J. E., … Milham, M.

P. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry, 19(6), 659–667. https://doi.org/10.1038/mp.2013.78

[102] Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., Breitner, J. C., Buckner, R. L., Calhoun, V. D., Castellanos, F. X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S. J., Courtney, W., Craddock, R. C., Di Martino, A., Dong, H. M., Fu, X., … Milham, M. P. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. Scientific data, 1, 140049. https://doi.org/10.1038/sdata.2014.49

[103] Provins C, MacNicol E, Seeley SH, Hagmann P and Esteban O (2023) Quality control in functional MRI studies with MRIQC and fMRIPrep. Front. Neuroimaging 1:1073734. doi: 10.3389/fnimg.2022.1073734

[104] Fallatah, S. M., Pizzini, F. B., Gomez-Anson, B., Magerkurth, J., De Vita, E., Bisdas, S., Jäger, H. R., Mutsaerts, H. J. M. M., & Golay, X. (2018). A visual quality control scale for clinical arterial spin labeling images. European radiology experimental, 2(1), 45. https://doi.org/10.1186/s41747-018-0073-2

[105] Pernet, Cyril. 2019. “SPM U+.” Open Science Framework. March 16, 2019. https://osf.io/wn3h8/.

[106] Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data 3, 160044 (2016).

[107] Dolui, S. et al. Structural Correlation-based Outlier Rejection (SCORE) algorithm for arterial spin labeling time series: SCORE: denoising algorithm for ASL. J. Magn. Reson. Imaging 45, 1786–1797 (2017).

[108] Dolui, S. SCRUB: a structural correlation and empirical robust Bayesian method for ASL data. In Proc. International Society for Magnetic Resonance in Medicine, Singapore, May http://archive.ismrm.org/2016/2880.html (ISMRM, 2016).

[109] Michael A. Chappell, Thomas F. Kirk, Martin S. Craig, Flora A. Kennedy McConnell, Moss Y. Zhao, Bradley J. MacIntosh, Thomas W. Okell, Mark W. Woolrich; BASIL: A toolbox for perfusion quantification using arterial spin labelling. Imaging Neuroscience 2023; 1 1–16. doi: https://doi.org/10.1162/imag_a_00041

[110] Li, W., Mo, W., Zhang, X., Squiers, J. J., Lu, Y., Sellke, E. W., Fan, W., DiMaio, J. M., & Thatcher, J. E. (2015). Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging. Journal of biomedical optics, 20(12), 121305. https://doi.org/10.1117/1.JBO.20.12.121305

[111] K. Li, C. Ye, Z. Yang, A. Carass, S. H. Ying, and J. L. Prince, “Quality assurance using outlier detection on an automatic segmentation method for the cerebellar peduncles,” in Medical Imaging 2016: Image Processing, vol. 9784. International Society for Optics and Photonics, 2016, p. 97841H.

[112] A. F. Mejia, M. B. Nebel, A. Eloyan, B. Caffo, and M. A. Lindquist, “Pca leverage: outlier detection for high-dimensional functional mag netic resonance imaging data,” Biostatistics, vol. 18, no. 3, pp. 521–536, 2017.

[113] C.-M. Kim, E. J. Hong, and R. C. Park, “Chest x-ray outlier detection model using dimension reduction and edge detection,” IEEE Access, 2021.

optics, vol. 20, no. 12, p. 121305, 2015.

[114] Ahmad, A., Parker, D., Dheer, S., Samani, Z. R., & Verma, R. (2023). 3D-QCNet - A pipeline for automated artifact detection in diffusion MRI images. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, 103, 102151. https://doi.org/10.1016/j.compmedimag.2022.102151

[115] Samani ZR, Alappatt JA, Parker D, Ismail AAO and Verma R (2020) QC-Automator: Deep Learning-Based Automated Quality Control for Diffusion MR Images. Front. Neurosci. 13:1456. doi: 10.3389/fnins.2019.01456

[116] Ding, Y., Suffren, S., Bellec, P., & Lodygensky, G. A. (2019). Supervised machine learning quality control for magnetic resonance artifacts in neonatal data sets. Human brain mapping, 40(4), 1290–1297. https://doi.org/10.1002/hbm.24449

[117] Hendriks, J., Mutsaerts, HJ., Joules, R. et al. A systematic review of (semi-)automatic quality control of T1-weighted MRI scans. Neuroradiology 66, 31–42 (2024). https://doi.org/10.1007/s00234-023-03256-0

[118] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9

[119] Diederik P. Kingma; Max Welling, An Introduction to Variational Autoencoders , now, 2019.

[120] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of theACM,63(11), 139–144.https://doi.org/10.1145/3422622

[121] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998, doi: 10.1109/5254.708428.

[122] Gao, L., Zhang, L., Liu, C., & Wu, S. (2020). Handling imbalanced medical image data: A deep-learning-based one-class classification approach. Artificial intelligence in medicine, 108, 101935. https://doi.org/10.1016/j.artmed.2020.101935

[123] Amari, S. I. (1993). Backpropagation and stochastic gradient descent method. Neurocomputing, 5(4-5), 185-196.

[124] Kingma, D. P. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[125] Block, H. D. (1962). The perceptron: A model for brain functioning. i. Reviews of Modern Physics, 34(1), 123.

[126] Agarap AF. Deep learning using rectifed linear units (relu). 2018. arXiv preprint arXiv:1803.08375.

[127] Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Für Med Phys 2019;29:102–27. https://doi.org/10.1016/j.zemedi.2018.11.002.

[128] Ronneberger O., Fischer P., and Brox T. U-Net: Convolutional networks for biomedical image segmentation. In MICCAI, 2015

[129] M R, Anala & Makker, Malika & Ashok, Aakanksha. (2019). Anomaly Detection in Surveillance Videos. 93-98. 10.1109/HiPCW.2019.00031.

[130] D. Kwon, K. Natarajan, S. C. Suh, H. Kim and J. Kim, "An Empirical Study on Network Anomaly Detection Using Convolutional Neural Networks," 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), Vienna, Austria, 2018, pp. 1595-1598, doi: 10.1109/ICDCS.2018.00178.

[131] Minhas, M. S., & Zelek, J. (2019). Anomaly detection in images. arXiv preprint arXiv:1905.13147.

[132] Q. Wei, Y. Ren, R. Hou, B. Shi, J. Y. Lo, and L. Carin, “Anomaly detection for medical images based on a one-class classification,” in Medical Imaging 2018: Computer-Aided Diagnosis, vol. 10575. International Society for Optics and Photonics, 2018, p. 105751M.

[133] Tschuchnig, M. E., & Gadermayr, M. (2022, March). Anomaly detection in medical imaging-a mini review. In Data Science–Analytics and Applications: Proceedings of the 4th International Data Science Conference–iDSC2021 (pp. 33-38). Wiesbaden: Springer Fachmedien Wiesbaden.

[134] Han, C., Rundo, L., Murao, K., Noguchi, T., Shimahara, Y., Milacski, Z. Á., ... & Satoh, S. I. (2021). MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. BMC bioinformatics, 22, 1-20.

[135] T. Schlegl, P. Seeb¨ ock, S. M. Waldstein, G. Langs, and U. Schmidt Erfurth, “f-anogan: Fast unsupervised anomaly detection with genera tive adversarial networks,” Medical image analysis, vol. 54, pp. 30–44, 2019.

[136] D. ˇ Stepec and D. Skoˇ caj, “Image synthesis as a pretext for unsupervised histopathological diagnosis,” in International Workshop on Simulation and Synthesis in Medical Imaging. Springer, 2020, pp. 174–183.

[137] T. Fujioka, K. Kubota, M. Mori, Y. Kikuchi, L. Katsuta, M. Kimura, E. Yamaga, M. Adachi, G. Oda, T. Nakagawa et al., “Efficient anomaly detection with generative adversarial network for breast ultrasound imaging,” Diagnostics, vol. 10, no. 7, p. 456, 2020.

[138] Peters R. (2006). Ageing and the brain. Postgraduate medical journal, 82(964), 84–88. https://doi.org/10.1136/pgmj.2005.036665

[139] VS Fonov, AC Evans, RC McKinstry, CR Almli and DL Collins, Unbiased nonlinear average age-appropriate brain templates from birth to adulthood, NeuroImage, Volume 47, Supplement 1, July 2009, Page S102 Organization for Human Brain Mapping 2009 Annual Meeting, DOI: http://dx.doi.org/10.1016/S1053-8119(09)70884-5

[140] Funck, T., Paquette, C., Evan s, A., & Thiel, A. (2014). Surface-based partial-volume correction for high-resolution PET. *NeuroImage, Suppl.Part 2, 102*, 674-87. doi:https://doi.org/10.1016/j.neuroimage.2014.08.037

[141] Blinkouskaya, Y., & Weickenmeier, J. (2021). Brain Shape Changes Associated With Cerebral Atrophy in Healthy Aging and Alzheimer's Disease. Frontiers in Mechanical Engineering, 7, 705653. https://doi.org/10.3389/fmech.2021.705653

[142] Dickstein, D. L., Kabaso, D., Rocher, A. B., Luebke, J. I., Wearne, S. L., & Hof, P. R. (2007). Changes in the structural complexity of the aged brain. Aging cell, 6(3), 275–284. https://doi.org/10.1111/j.1474-9726.2007.00289.x

[143] Sweeney, M. D., Kisler, K., Montagne, A., Toga, A. W., & Zlokovic, B. V. (2018). The role of brain vasculature in neurodegenerative disorders. Nature neuroscience, 21(10), 1318–1331. https://doi.org/10.1038/s41593-018-0234-x

[144] Payne, S. (2016). “Cerebral Autoregulation : Control of Blood Flow in the Brain,” in SpringerBriefs in Bioengineering, 1st Edn, (Cham: Springer International Publishing).

[145] Huber, W., Zanner, R., Schneider, G., Schmid, R., & Lahmer, T. (2019). Assessment of Regional Perfusion and Organ Function: Less and Non-invasive Techniques. *Frontiers in medicine*, *6*, 50. https://doi.org/10.3389/fmed.2019.00050

[146] Fantini, S., Sassaroli, A., Tgavalekos, K. T., & Kornbluth, J. (2016). Cerebral blood flow and autoregulation: current measurement techniques and prospects for noninvasive optical methods. *Neurophotonics*, *3*(3), 031411. https://doi.org/10.1117/1.NPh.3.3.031411

[147] Claassen, J. A. H. R., Thijssen, D. H. J., Panerai, R. B., & Faraci, F. M. (2021). Regulation of cerebral blood flow in humans: physiology and clinical implications of autoregulation. *Physiological reviews*, *101*(4), 1487–1559. https://doi.org/10.1152/physrev.00022.2020

[148] Zhang, N., Gordon, M. L., Ma, Y., Chi, B., Gomar, J. J., Peng, S., Kingsley, P. B., Eidelberg, D., & Goldberg, T. E. (2018). The Age-Related Perfusion Pattern Measured With Arterial Spin Labeling MRI in Healthy Subjects. Frontiers in aging neuroscience, 10, 214. https://doi.org/10.3389/fnagi.2018.00214

[149] Hernandez-Garcia, L., Lahiri, A., & Schollenberger, J. (2019). Recent progress in ASL. NeuroImage, 187, 3–16. https://doi.org/10.1016/j.neuroimage.2017.12.095

[150] Chen, J. J., Rosas, H. D., & Salat, D. H. (2011). Age-associated reductions in cerebral blood flow are independent from regional atrophy. NeuroImage, 55(2), 468–478. https://doi.org/10.1016/j.neuroimage.2010.12.032

[151] Bertsch, K., Hagemann, D., Hermes, M., Walter, C., Khan, R., & Naumann, E. (2009). Resting cerebral blood flow, attention, and aging. Brain research, 1267, 77–88. https://doi.org/10.1016/j.brainres.2009.02.053

[152] Tarumi, T., Ayaz Khan, M., Liu, J., Tseng, B. Y., Parker, R., Riley, J., Tinajero, C., & Zhang, R. (2014). Cerebral hemodynamics in normal aging: central artery stiffness, wave reflection, and pressure pulsatility. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism, 34(6), 971–978. https://doi.org/10.1038/jcbfm.2014.44

[153] H. W. Müller-Gärtner et al., (1992). “Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects,” J. Cerebral Blood Flow Metabolism, vol. 12, no. 4, pp. 571–583, Jul.

[154] Fischl B. (2012). FreeSurfer. NeuroImage, 62(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

[155] Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., Jenkinson, M., & WU-Minn HCP Consortium (2013). The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage, 80, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

[156] Greve, D. N., Svarer, C., Fisher, P. M., Feng, L., Hansen, A. E., Baare, W., Rosen, B., Fischl, B., & Knudsen, G. M. (2014). Cortical surface-based analysis reduces bias and variance in kinetic modeling of brain PET data. NeuroImage, 92, 225–236. https://doi.org/10.1016/j.neuroimage.2013.12.021

[157] Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage, 9(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

[158] Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences of the United States of America, 97(20), 11050–11055. https://doi.org/10.1073/pnas.200033797

[159] Park, H. J., Lee, J. D., Chun, J. W., Seok, J. H., Yun, M., Oh, M. K., & Kim, J. J. (2006). Cortical surface-based analysis of 18F-FDG PET: measured metabolic abnormalities in schizophrenia are affected by cortical structural

abnormalities. NeuroImage, 31(4), 1434–1444. https://doi.org/10.1016/j.neuroimage.2006.02.001

[160] Ferreira D, Nedelska Z, Graff-Radford J, Przybelski SA, Lesnick TG, Schwarz CG, Botha H, Senjem ML, Fields JA, Knopman DS, Savica R, Ferman TJ, Graff-Radford NR, Lowe VJ, Jack CR, Petersen RC, Lemstra AW, van de Beek M, Barkhof F, Blanc F, Loureiro de Sousa P, Philippi N, Cretin B, Demuynck C, Hort J, Oppedal K, Boeve BF, Aarsland D, Westman E, Kantarci K. Cerebrovascular disease, neurodegeneration, and clinical phenotype in dementia with Lewy bodies. Neurobiol Aging. 2021 Sep;105:252-261. doi: 10.1016/j.neurobiolaging.2021.04.029. Epub 2021 May 14. PMID: 34130107; PMCID: PMC8338792.

[161] Taso, M., Munsch, F., Zhao, L., & Alsop, D. C. (2021). Regional and depth-dependence of cortical blood-flow assessed with high-resolution Arterial Spin Labeling (ASL). Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism, 41(8), 1899–1911. https://doi.org/10.1177/0271678X20982382

[162] Verclytte, S., Lopes, R., Delmaire, C., Ferre, J. C., Pasquier, F., & Leclerc, X. (2015). Optimization of brain perfusion image quality by cortical surface-based projection of arterial spin labeling maps in early-onset Alzheimer's disease patients. European radiology, 25(8), 2479–2484. https://doi.org/10.1007/s00330-015-3652-0

[163] Leidhin, C. N., McMorrow, J., Carey, D., Newman, L., Williamson, W., Fagan, A. J., Chappell, M. A., Kenny, R. A., Meaney, J. F., & Knight, S. P. (2021). Age-related normative changes in cerebral perfusion: Data from The Irish Longitudinal Study on Ageing (TILDA). NeuroImage, 229, 117741. https://doi.org/10.1016/j.neuroimage.2021.117741

[164] Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage, 23 Suppl 1, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

[165] Kirk, T. F., Coalson, T. S., Craig, M. S., & Chappell, M. A. (2020). Toblerone: Surface-Based Partial Volume Estimation. IEEE transactions on medical imaging, 39(5), 1501–1510. https://doi.org/10.1109/TMI.2019.2951080

[166] Kearney, P. M., Cronin, H., O'Regan, C., Kamiya, Y., Savva, G. M., Whelan, B., & Kenny, R. (2011). Cohort profile: the Irish Longitudinal Study on Ageing. International journal of epidemiology, 40(4), 877–884. https://doi.org/10.1093/ije/dyr116

[167] Whelan, B. J., & Savva, G. M. (2013). Design and methodology of the Irish Longitudinal Study on Ageing. Journal of the American Geriatrics Society, 61 Suppl 2, S265–S268. https://doi.org/10.1111/jgs.12199

[168] Patenaude, B., Smith, S. M., Kennedy, D. N., & Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage, 56(3), 907–922. https://doi.org/10.1016/j.neuroimage.2011.02.046

[169] Glasser, M. F., & Van Essen, D. C. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. The Journal of neuroscience : the official journal of the Society for Neuroscience, 31(32), 11597–11616. https://doi.org/10.1523/JNEUROSCI.2180-11.2011

[170] Buxton R. B. (2005). Quantifying CBF with arterial spin labeling. Journal of magnetic resonance imaging : JMRI, 22(6), 723–726. https://doi.org/10.1002/jmri.20462

[171] Griffanti, L., Raman, B., Alfaro-Almagro, F., Filippini, N., Cassar, M. P., Sheerin, F., Okell, T. W., Kennedy McConnell, F. A., Chappell, M. A., Wang, C., Arthofer, C., Lange, F. J., Andersson, J., Mackay, C. E., Tunnicliffe, E. M., Rowland, M., Neubauer, S., Miller, K. L., Jezzard, P., & Smith, S. M. (2021). Adapting the UK Biobank Brain Imaging Protocol and Analysis Pipeline for the C-MORE Multi-Organ Study of COVID-19 Survivors. Frontiers in neurology, 12, 753284. https://doi.org/10.3389/fneur.2021.753284

[172] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage, 31(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

[173] Jefferson, A. L., Cambronero, F. E., Liu, D., Moore, E. E., Neal, J. E., Terry, J. G., Nair, S., Pechman, K. R., Rane, S., Davis, L. T., Gifford, K. A., Hohman, T. J., Bell, S. P., Wang, T. J., Beckman, J. A., & Carr, J. J. (2018). Higher Aortic Stiffness Is Related to Lower Cerebral Blood Flow and Preserved Cerebrovascular Reactivity in Older Adults. Circulation, 138(18), 1951–1962. https://doi.org/10.1161/CIRCULATIONAHA.118.032410

[174] Biagi, L., Abbruzzese, A., Bianchi, M. C., Alsop, D. C., Del Guerra, A., & Tosetti, M. (2007). Age dependence of cerebral perfusion assessed by magnetic resonance continuous arterial spin labeling. Journal of magnetic resonance imaging : JMRI, 25(4), 696–702. https://doi.org/10.1002/jmri.20839

[175] Zhao, M. Y., Mezue, M., Segerdahl, A. R., Okell, T. W., Tracey, I., Xiao, Y., & Chappell, M. A. (2017). A systematic study of the sensitivity of partial volume correction methods for the quantification of perfusion from pseudo-continuous arterial spin labeling MRI. NeuroImage, 162, 384–397. https://doi.org/10.1016/j.neuroimage.2017.08.072

[176] Meltzer, C. C., Cantwell, M. N., Greer, P. J., Ben-Eliezer, D., Smith, G., Frank, G., Kaye, W. H., Houck, P. R., & Price, J. C. (2000). Does cerebral blood flow decline in healthy aging? A PET study with partial-volume

correction. Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 41(11), 1842–1848.

[177] Inoue, K., Ito, H., Goto, R., Nakagawa, M., Kinomura, S., Sato, T., Sato, K., & Fukuda, H. (2005). Apparent CBF decrease with normal aging due to partial volume effects: MR-based partial volume correction on CBF SPECT. Annals of nuclear medicine, 19(4), 283–290. https://doi.org/10.1007/BF02984620

[178] De Vis, J. B., Hendrikse, J., Bhogal, A., Adams, A., Kappelle, L. J., & Petersen, E. T. (2015). Age-related changes in brain hemodynamics; A calibrated MRI study. Human brain mapping, 36(10), 3973–3987. https://doi.org/10.1002/hbm.22891

[179] Liu, Y., Zhu, X., Feinberg, D., Guenther, M., Gregori, J., Weiner, M. W., & Schuff, N. (2012). Arterial spin labeling MRI study of age and gender effects on brain perfusion hemodynamics. Magnetic resonance in medicine, 68(3), 912–922. https://doi.org/10.1002/mrm.23286

[180] Lu, H., Xu, F., Rodrigue, K. M., Kennedy, K. M., Cheng, Y., Flicker, B., Hebrank, A. C., Uh, J., & Park, D. C. (2011). Alterations in cerebral metabolic rate and blood supply across the adult lifespan. Cerebral cortex (New York, N.Y. : 1991), 21(6), 1426–1434. https://doi.org/10.1093/cercor/bhq224

[181] Parkes, L. M., Rashid, W., Chard, D. T., & Tofts, P. S. (2004). Normal cerebral perfusion measurements using arterial spin labeling: reproducibility, stability, and age and gender effects. Magnetic resonance in medicine, 51(4), 736–743. https://doi.org/10.1002/mrm.20023

[182] Martin, A. J., Friston, K. J., Colebatch, J. G., & Frackowiak, R. S. (1991). Decreases in regional cerebral blood flow with normal aging. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism, 11(4), 684–689. https://doi.org/10.1038/jcbfm.1991.121

[183] Fierstra, J., Poublanc, J., Han, J. S., Silver, F., Tymianski, M., Crawley, A. P., Fisher, J. A., & Mikulis, D. J. (2010). Steal physiology is spatially associated with cortical thinning. Journal of neurology, neurosurgery, and psychiatry, 81(3), 290–293. https://doi.org/10.1136/jnnp.2009.188078

[184] Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S., Busa, E., Morris, J. C., Dale, A. M., & Fischl, B. (2004). Thinning of the cerebral cortex in aging. Cerebral cortex (New York, N.Y. : 1991), 14(7), 721–730. https://doi.org/10.1093/cercor/bhh032

[185] Bangen KJ, Werhane ML, Weigand AJ, Edmonds EC, Delano-Wood L, Thomas KR, Nation DA, Evangelista ND, Clark AL, Liu TT, Bondi MW. Reduced Regional Cerebral Blood Flow Relates to Poorer Cognition in Older Adults With Type 2 Diabetes. Front Aging Neurosci. 2018 Sep 10;10:270. doi: 10.3389/fnagi.2018.00270. PMID: 30250430; PMCID: PMC6139361.

[186] Alisch, J. S. R., Khattar, N., Kim, R. W., Cortina, L. E., Rejimon, A. C., Qian, W., Ferrucci, L., Resnick, S. M., Spencer, R. G., & Bouhrara, M. (2021). Sex and age-related differences in cerebral blood flow investigated using pseudo-continuous arterial spin labeling magnetic resonance imaging. Aging, 13(4), 4911–4925. https://doi.org/10.18632/aging.202673

[187] Han, H., Lin, Z., Soldan, A., Pettigrew, C., Betz, J. F., Oishi, K., Li, Y., Liu, P., Albert, M., & Lu, H. (2022). Longitudinal Changes in Global Cerebral Blood Flow in Cognitively Normal Older Adults: A Phase-Contrast MRI Study. Journal of magnetic resonance imaging : JMRI, 56(5), 1538–1545. https://doi.org/10.1002/jmri.28133

[188] Campbell, A. M., & Beaulieu, C. (2006). Pulsed arterial spin labeling parameter optimization for an elderly population. Journal of magnetic resonance imagin : JMRI, 23(3), 398–403. https://doi.org/10.1002/jmri.20503

[189] Borogovac A, Asllani I (2012) Arterial spin labeling (ASL) fMRI: advantages, theoretical constrains and experimental challenges in neurosciences. Int J Biomed Imaging 2012:818456

[190] Bastiani M, Cottaar M, Fitzgibbon SP, Suri S, Alfaro-Almagro F, Sotiropoulos SN, Jbabdi S, Andersson JLR. Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. Neuroimage. 2019 Jan 1;184:801-812. doi: 10.1016/j.neuroimage.2018.09.073. Epub 2018 Sep 26. PMID: 30267859; PMCID: PMC6264528.

[191] Restom, K., Behzadi, Y., & Liu, T. T. (2006). Physiological noise reduction for arterial spin labeling functional MRI. *NeuroImage*, *31*(3), 1104–1115. https://doi.org/10.1016/j.neuroimage.2006.01.026

[192] Hendrikse J, Petersen ET, Golay X (2012) Vascular disorders: insights from arterial spin labeling. Neuroimaging Clin N Am 22:259–269

[193] Alsop DC, Dai W, Grossman M, Detre JA (2010) Arterial spin labeling blood flow MRI: its role in the early characterization of Alzheimer's disease. J Alzheimers Dis 20:871–880

[194] Detre JA, Rao H, Wang DJ, Chen YF, Wang Z (2012) Applications of arterial spin labeled MRI in the brain. J Magn Reson Imaging 35:1026–1037

[195] Alsop DC, Detre JA (1996) Reduced transit-time sensitivity in noninvasive magnetic resonance imaging of human cerebral blood flow. J Cereb Blood Flow Metab 16:1236–1249

[196] Fernández-Seara M, Edlow BL, Hoang A, Wang J, Feinberg DA, Detre JA (2008) Minimizing acquisition time of arterial spin labeling at 3T. Magn Reson Med 59:1467–1471

[197] Li, Y., Liu, P., Li, Y., Fan, H., Su, P., Peng, S. L., Park, D. C., Rodrigue, K. M., Jiang, H., Faria, A. V., Ceritoglu, C., Miller, M., Mori, S., & Lu, H. (2019). ASL-MRICloud: An online tool for the processing of ASL MRI data. *NMR in biomedicine*, *32*(2), e4051. https://doi.org/10.1002/nbm.4051

[198] Physimals. OXASL. GitHub. https://github.com/physimals/oxasl

[199] Razlighi, Q.R, Kehtarnavaz, N, and Yousefi, S. "Evaluating Similarity Measures for Brain Image Registration." Journal of Visual Communication and Image Representation 24.7 (2013): 977-87. Web.

[200] Chen T, Yuan M, Tang J, Lu L. Digital Analysis of Smart Registration Methods for Magnetic Resonance Images in Public Healthcare. Front Public Health. 2022 Jun 6;10:896967. doi: 10.3389/fpubh.2022.896967. PMID: 35734757; PMCID: PMC9207932.

[201] Castillo, R., Castillo, E., Fuentes, D., Ahmad, M., Wood, A. M., Ludwig, M. S., & Guerrero, T. (2013). A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Physics in medicine and biology, 58(9), 2861–2877. https://doi.org/10.1088/0031-9155/58/9/2861

[202] Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging. 2003; vol. 22(no. 8):986–1004. [PubMed: 12906253]

[203] Gholipour A, Kehtarnavaz N, Briggs R, Devous M, Gopinath K. Brain functional localization: a survey of image registration techniques. IEEE Transactions on Medical Imaging. 2007; vol. 26(no. 4):427–451. [PubMed: 17427731]

[204] Moore CS, Liney GP, Beavis AW. Quality assurance of registration of CT and MRI data sets for treatment planning of radiotherapy for head and neck cancers. J Appl Clin Med Phys. 2004 Winter;5(1):25-35. doi: 10.1120/jacmp.v5i1.1951. Epub 2004 Jan 1. PMID: 15753931; PMCID: PMC5723443.

[205] J. P. W. Pluim, S. E. A. Muenzing, K. A. J. Eppenhof and K. Murphy, "The truth is hard to make: Validation of medical image registration," 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 2016, pp. 2294-2300, doi: 10.1109/ICPR.2016.7899978.

[206] Joshua Bierbrier, Houssem-Eddine Gueziri, D. Louis Collins, Estimating medical image registration error and confidence: A taxonomy and scoping review, Medical Image Analysis, Volume 81, 2022, 102531, ISSN 1361-8415, https://doi.org/10.1016/j.media.2022.102531.

[207] W. R. Crum, O. Camara, and D. L. G. Hill, “Generalized overlap measures for evaluation and validation in medical image analysis,” IEEE Trans Med Imaging, vol. 25, no. 11, pp. 1451–1461, 2006

[208] T. Netsch, P. Rsch, J. Weese, A. van Muiswinkel, and P. Desmedt, “Grey value-based 3-D registration of functional MRI time-series: comparison of interpolation order and similarity measure,” presented at the Medical Image Computing and Computer Assisted Intervention - MICCAI, 2000, vol. 3979, pp. 1148–1159.

[209] Xi Cheng, Li Zhang & Yefeng Zheng (2018) Deep similarity learning for multimodal medical images, Computer Methods in Biomechanics and

Biomedical Engineering: Imaging & Visualization, 6:3, 248-252, DOI: 10.1080/21681163.2015.1135299

[210] Steffen Czolbe, Paraskevas Pegios, Oswin Krause, Aasa Feragen, Semantic similarity metrics for image registration, Medical Image Analysis, Volume 87, 2023, 102830, ISSN 1361-8415, https://doi.org/10.1016/j.media.2023.102830.

[211] Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

[212] Paschoal, A. M., Woods, J. G., Pinto, J., Bron, E. E., Petr, J., Kennedy McConnell, F. A., Bell, L., Dounavi, M. E., van Praag, C. G., Mutsaerts, H. J. M. M., Taylor, A. O., Zhao, M. Y., Brumer, I., Chan, W. S. M., Toner, J., Hu, J., Zhang, L. X., Domingos, C., Monteiro, S. P., Figueiredo, P., … Anazodo, U. (2024). Reproducibility of arterial spin labeling cerebral blood flow image processing: A report of the ISMRM open science initiative for perfusion imaging (OSIPI) and the ASL MRI challenge. Magnetic resonance in medicine, 92(2), 836–852. https://doi.org/10.1002/mrm.30081

[213] Fan, H., Mutsaerts, H. J. M. M., Anazodo, U., Arteaga, D., Baas, K. P. A., Buchanan, C., Camargo, A., Keil, V. C., Lin, Z., Lindner, T., Hirschler, L., Hu, J., Padrela, B. E., Taghvaei, M., Thomas, D. L., Dolui, S., & Petr, J. (2024). ISMRM Open Science Initiative for Perfusion Imaging (OSIPI): ASL pipeline inventory. Magnetic resonance in medicine, 91(5), 1787–1802. https://doi.org/10.1002/mrm.29869

[214] Ma, Z., Reich, D. S., Dembling, S., Duyn, J. H., & Koretsky, A. P. (2022). Outlier detection in multimodal MRI identifies rare individual phenotypes among more than 15,000 brains. *Human brain mapping*, *43*(5), 1766–1782. https://doi.org/10.1002/hbm.25756

[215] Pollock, J. M., Tan, H., Kraft, R. A., Whitlow, C. T., Burdette, J. H., & Maldjian, J. A. (2009). Arterial spin-labeled MR perfusion imaging: clinical applications. *Magnetic resonance imaging clinics of North America*, *17*(2), 315–338. https://doi.org/10.1016/j.mric.2009.01.008

[216] Gudbjartsson, H., & Patz, S. (1995). The Rician distribution of noisy MRI data. *Magnetic resonance in medicine*, *34*(6), 910-914.

[217] Dietrich, O., Raya, J. G., Reeder, S. B., Reiser, M. F., & Schoenberg, S. O. (2007). 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*, *26*(2), 375-385.

[218] Reeder, S. B., Wintersperger, B. J., Dietrich, O., Lanz, T., Greiser, A., Reiser, M. F., ... & Schoenberg, S. O. (2005). Practical approaches to the evaluation of signal-to-noise ratio performance with parallel imaging: application with cardiac imaging and a 32-channel cardiac coil. *Magnetic*

Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, *54*(3), 748-754.

[219] Petersen, E. T., Lim, T., & Golay, X. (2006). Model-free arterial spin labeling quantification approach for perfusion MRI. *Magnetic resonance in medicine*, *55*(2), 219–232. https://doi.org/10.1002/mrm.20784

[220] Dominik Weishaupt, Victor D. Kochli, Borut Marincek, How does MRI work?, Springer Verlag Berlin Heidelberg New York, 2003.

[221] Shirzadi, Z., Stefanovic, B., Chappell, M. A., Ramirez, J., Schwindt, G., Masellis, M., Black, S. E., & MacIntosh, B. J. (2018). Enhancement of automated blood flow estimates (ENABLE) from arterial spin-labeled MRI. Journal of magnetic resonance imaging : JMRI, 47(3), 647–655. https://doi.org/10.1002/jmri.25807

[222] Mercadante AA, Tadi P. Neuroanatomy, Gray Matter. [Updated 2023 Jul 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK553239/

[223] Van Osch MJ, Teeuwisse WM, van Walderveen MA, Hendrikse J, Kies DA et al. (2009) Can arterial spin labeling detect white matter perfusion signal? Magn Reson Med 62: 165-173. doi:10.1002/mrm.22002. PubMed: 19365865.

[224] Chalela JA, Alsop DC, Gonzalez-Atavales JB, et al. Magnetic resonance perfusion imaging in acute ischemic stroke using continuous arterial spin labeling. Stroke 2000; 31: 680–687.

[225] Donahue J, Sumer S and Wintermark M. Assessment of collateral flow in patients with cerebrovascular disorders. J Neuroradiol 2014; 41: 234–242.

[226] MacIntosh BJ and Graham SJ. Magnetic resonance ima ging to visualize stroke and characterize stroke recovery: a review. Front Neurol 2013; 4: 60.

[227] Zaharchuk G. Arterial spin label imaging of acute ische mic stroke and transient ischemic attack. Neuroimaging Clin N Am 2011; 21: 285–301.

[228] Mutsaerts, H. J., Petr, J., Václavů, L., van Dalen, J. W., Robertson, A. D., Caan, M. W., Masellis, M., Nederveen, A. J., Richard, E., & MacIntosh, B. J. (2017). The spatial coefficient of variation in arterial spin labeling cerebral blood flow images. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, *37*(9), 3184–3192. https://doi.org/10.1177/0271678X16683690

[229] Wu, W. C., Lin, S. C., Wang, D. J., Chen, K. L., & Li, Y. D. (2013). Measurement of cerebral white matter perfusion using pseudocontinuous arterial spin labeling 3T magnetic resonance imaging--an experimental and theoretical investigation of feasibility. PloS one, 8(12), e82679. https://doi.org/10.1371/journal.pone.0082679

[230] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas,

J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[231] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium (2013). The WU-Minn Human Connectome Project: an overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

[232] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful represen tations in a deep network with a local denoising criterion." Journal of machine learning research, vol. 11, no. 12, 2010.

[233] J. Sun, X. Wang, N. Xiong, and J. Shao, "Learning sparse representation with variational auto-encoder for anomaly detection," pp. 33353 33361, 2018.

[234] Rezende, D., & Mohamed, S. (2015, June). Variational inference with normalizing flows. In International conference on machine learning (pp. 1530-1538). PMLR.

[235] Larsen, A. B. L., Sønderby, S. K., Larochelle, H., & Winther, O. (2016, June). Autoencoding beyond pixels using a learned similarity metric. In International conference on machine learning (pp. 1558-1566). PMLR.

[236] Chen, Y. C. (2017). A tutorial on kernel density estimation and recent advances. Biostatistics & Epidemiology, 1(1), 161-187.

[237] A. R. Lahitani, A. E. Permanasari and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," 2016 4th International Conference on Cyber and IT Service Management, Bandung, Indonesia, 2016, pp. 1-6, doi: 10.1109/CITSM.2016.7577578.

[238] Nahm F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. Korean journal of anesthesiology, 75(1), 25–36. https://doi.org/10.4097/kja.21209

[239] J. An, S. Cho, Variational autoencoder based anomaly detection using reconstruction probability, Tech. rep., SNU Data Mining Center (2015).

[240] D. Park, Y. Hoshi, C. C. Kemp, A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational au toencoder, IEEE Robotics and Automation Letters 3 (3) (2018) 1544–1551 (2018).

[241] Wang X., Du Y., Lin S., Cui P., Shen Y., Yang Y., adVAE: A self-adversarial variational autoencoder with Gaussian anomaly prior knowledge for anomaly detection, Knowledge-Based Systems 190 (2020).

[242] Hu J, Craig MS, Knight SP, De Looze C, Meaney JF, Kenny RA, Chen X, Chappell MA. Regional changes in cerebral perfusion with age when accounting for changes in gray-matter volume. Magn Reson Med. 2025 Apr;93(4):1807-1820. doi: 10.1002/mrm.30376. Epub 2024 Nov 20. PMID: 39568213; PMCID: PMC11782718.

[243] Y. Zhang, J. -Q. Zheng and M. Chappell, "Bayesian Inference for Non-Linear Forward Model by Using a VAE-Based Neural Network Structure," in IEEE Transactions on Signal Processing, vol. 72, pp. 1400-1411, 2024, doi: 10.1109/TSP.2024.3374115.

[244] Sewak, M., Sahay, S. K., & Rathore, H. (2020). An overview of deep learning architecture of deep neural networks and autoencoders. Journal of Computational and Theoretical Nanoscience, 17(1), 182-188.

[245] Henkelman, R. M., Stanisz, G. J., & Graham, S. J. (2001). Magnetization transfer in MRI: a review. *NMR in biomedicine*, *14*(2), 57–64. https://doi.org/10.1002/nbm.683

[246] Camille Van Assel, Gabriel Mangeat, Benjamin De Leener, Nikola Stikov, Caterina Mainero, and Julien Cohen-Adad. Partial volume effect correction for surface-based cortical mapping (https://ismrm.gitlab.io/2017/5068.html). ISMRM.2017.