Synthetic Correlated Diffusion Imaging for Prostate Cancer Detection and Risk Assessment

by

Hayden Gunraj

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2023

Author's Declaration

This thesis consists of material which I authored or co-authored (see Statement of Contributions on the following page). This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The following paper is included in Chapter 3 of this thesis:

A. Wong, H. Gunraj, V. Sivan, and M. A. Haider. Synthetic correlated diffusion imaging hyperintensity delineates clinically significant prostate cancer. *Scientific Reports*, 12:3376, 2022.

I was a co-author with major contributions to the study design and data analysis, as well as writing and editing of the paper. In particular, I implemented and performed all software-based experiments and wrote much of the methods, results, and discussion. A. Wong formulated synthetic correlated diffusion imaging and conceived the study, while V. Sivan conducted initial data exploration. M. A. Haider reviewed the data and provided clinical interpretation of images.

The included material is a paraphrased adaptation of the paper, and updated methods and results are presented in Chapter 3 which were not included in the paper.

Abstract

Prostate cancer (PCa) is the second most common form of cancer among men worldwide and the most frequently diagnosed cancer among men in 112 countries. While the overall 5-year survival rate for prostate cancer is very high, prognosis varies considerably depending on how early PCa is diagnosed and how aggressively it develops. As such, PCa screening is critical for early detection and treatment of PCa. However, many PCas develop slowly and pose a minimal risk of PCa-related mortality, in which case treatment can be limited to active surveillance of tumour development.

Over the last few decades, magnetic resonance imaging (MRI) been used extensively for PCa screening and assessment. In particular, multi-parametric magnetic resonance imaging (mpMRI), where multiple MRI modalities are acquired, is commonly used for PCa imaging. However, the use of mpMRI requires radiologists to interpret multiple MRI images in parallel, resulting in increased inter-observer variability. This is especially true forradiologists with less experience interpreting prostate MRI images. In an effort to address these concerns, a computational MRI modality known as correlated diffusion imaging (CDI) was introduced, with initial results showing promise for CDI as a PCa screening tool. However, CDI is uncalibrated and strongly dependent on the underlying MRI protocols used to compute it, which leads to inconsistencies across different protocols and significant inter- and intra-patient variability.

In this thesis, a computational MRI technique known as synthetic correlated diffusion imaging (CDI^{s}) is introduced. CDI^{s} extends CDI through the addition of synthetic DWI and per-patient calibration, thereby providing flexibility and consistency beyond that of CDI. Additionally, a gradient-based optimization framework is developed through which the parameters of CDI^{s} may be optimized for downstream clinical tasks.

The proposed CDI^{s} and optimization framework were evaluated against current standard MRI modalities using a clinical MRI dataset comprising 200 PCa patients. Through clinical interpretation by an experienced radiologist, CDI^{s} was found to provide better tissue contrast between healthy, low-risk PCa, and high-risk PCa than standard MRI modalities. This suggests that CDI^{s} provides visual indications of PCa and PCa risk level, which may allow radiologists to make more accurate and consistent conclusions from imaging alone. CDI^{s} may also be used to guide prostate biopsies, potentially indicating better biopsy locations and reducing the number of biopsies required.

Upon quantitative evaluation, CDI^{s} achieved a voxel-level area under the receiver operating characteristic curve (AUC) of 0.8446 for separation of healthy and PCa tissue, representing an increase of 0.0315 ($p < 0.0001$) over the best-performing standard MRI

modality and indicating the potential of CDI^s for PCa screening and diagnosis. Moreover, CDI^s achieved a voxel-level AUC of 0.8530 for distinguishing between high- and low-risk cancers, representing an increase of 0.1590 ($p < 0.0001$) over the best-performing standard MRI modality and indicating the potential of CDI^s for PCa risk assessment.These results suggest that CDI^s may improve voxel-level identification of PCa, which is valuable for PCa localization and segmentation. Moreover, machine learning models trained on CDI^s images can benefit from this improved voxel-level contrast, potentially achieving better diagnostic, prognostic, or segmentation performance than models trained on standard MRI images.

Acknowledgements

I would like to thank my supervisor, Prof. Alexander Wong, for his constant encouragement and excitement about new ideas. The freedom you've given me to pursue a variety of projects with often unrelated applications has given me a wide breadth of knowledge and experience that will serve me for years to come. More importantly, thank you for your kindness, support, and willingness to consider my opinions even when they conflicted with your own. It's easy to enjoy research with a supervisor like you.

I would like to thank Prof. Stewart McLachlin and Prof. John Zelek for reviewing my thesis and providing valuable suggestions to better explain and contextualize my research. I appreciate you taking the time to help make my thesis the best that it can be. I would also like to thank Dr. Masoom Haider, MD for his collaboration on this research. Your insights into the clinical potential and limitations of this work have helped motivate and validate my research directions.

I would be remiss not to mention Prof. Nilesh Ghugre and Venkat Swaminathan, who were my supervisors during my undergraduate co-op term at Sunnybrook Research Institute. This thesis may not have come to be if you had not piqued my interest in both research and MRI early in my academic career, and it was a pleasure to learn about MR images and image processing under your supervision.

I would also like to thank my friends and colleagues in the Vision and Image Processing Lab for their support, encouragement, and ideas throughout my time in the lab. I will miss our chats about research, careers, and life in general, as well as our board game nights. Special thanks to Vignesh Sivan for helping me get familiar with diffusion-weighted imaging when I began this research, and to Amy Tai for constantly listening to (unsolicited) explanations of my work and offering ideas and suggestions.

My family and friends have supported me throughout my time at the University of Waterloo. My parents in particular have been immensely supportive, and have always encouraged me to seek out new and exciting opportunities and learn as much as I can. So many of my friends have supported me throughout this degree, and although I cannot list all of them I would especially like to thank Chris Tran, Bob Long, Mandy Coleman, Cindy Chin, Elaine Au-Yang, Emma Tung, Victoria Huang, and Sam Wong. I would also like to offer a blanket apology to all of my friends for repeatedly trying to introduce and explain technical topics (especially MRI) in unrelated conversations.

Finally, I would like to thank the Natural Sciences and Engineering Research Council of Canada and the Government of Ontario for funding this research.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADC	apparent diffusion coefficient
AS	anterior fibromuscular stroma
AUC	area under the receiver operating characteristic curve
BPH	benign prostatic hyperplasia
CDI	correlated diffusion imaging
CDI^s	synthetic correlated diffusion imaging
CHB-DWI	computed high b-value diffusion-weighted imaging
CI	confidence interval
csPCa	clinically significant prostate cancer
DCE	dynamic contrast-enhanced
DDC	distributed diffusion coefficient
DFOV	display field of view
DW	diffusion-weighted
DWI	diffusion-weighted imaging
GG	grade group
GS	Gleason score
insPCa	clinically insignificant prostate cancer

ISUP	International Society of Urological Pathology
ME	monoexponential
MEM	monoexponential model
mpMRI	multi-parametric magnetic resonance imaging
MR	magnetic resonance
MRI	magnetic resonance imaging
NCCN	National Comprehensive Cancer Network
PCa	prostate cancer
PDF	probability density function
PET	positron emission tomography
PI-RADS	Prostate Imaging–Reporting and Data System
PSA	prostate-specific antigen
PZ	peripheral zone
ROC	receiver operating characteristic
SE	stretched exponential
SEM	stretched exponential model
SGD	stochastic gradient descent
SI	signal intensity
SNR	signal-to-noise ratio
T2w	T2-weighted
TRUS	transrectal ultrasound
TZ	transition zone
WHO	World Health Organization
WMW	Wilcoxon-Mann-Whitney

List of Symbols

α	Water molecular diffusion heterogeneity
$\hat{\alpha}$	Estimated water molecular diffusion heterogeneity
$\hat{\alpha}$	Estimated water molecular diffusion heterogeneity
γ	Proton gyromagnetic ratio
δ	Gradient pulse duration in diffusion-weighted imaging
Δ	Time between gradient pulses in diffusion-weighted imaging
$\boldsymbol{\theta}$	Vector of parameters of a signal decay model in diffusion-weighted imaging
$\hat{\boldsymbol{\theta}}(\mathbf{x})$	Vector of estimated parameters of a signal decay model at spatial location $\mathbf{x}$ in diffusion-weighted imaging
$\boldsymbol{\rho}$	N-vector of exponents for different diffusion-weighted acquisitions
$\sigma_\beta(\cdot)$	Sigmoid function with parameter β
σ_I	Intensity standard deviation in a bilateral filter function
Σ	Covariance matrix of a Gaussian distribution
A	Apparent diffusion coefficient
$\hat{A}$	Estimated apparent diffusion coefficient
b, b_i, b_a, b_b	b-value in diffusion-weighted imaging
$\mathbf{b}$	N-vector of b-values for different diffusion-weighted acquisitions
D	Distributed diffusion coefficient
$\hat{D}$	Estimated distributed diffusion coefficient

$f(\mathbf{x}')$	Prior probability density function in synthetic correlated diffusion imaging
$f_b(\mathbf{x}'; \mathbf{x}, \Sigma, \sigma_I)$	3-dimensional bilateral filter in synthetic correlated diffusion imaging
$f_g(\mathbf{x}'; \mathbf{x}, \Sigma)$	3-dimensional Gaussian probability density function in synthetic correlated diffusion imaging
G	Gradient pulse strength in diffusion-weighted imaging
$H(\cdot)$	Heaviside step function
$I(b_a, b_b; \boldsymbol{\theta})$	Weighted product integral of synthetic diffusion-weighted signals over interval $[b_a, b_b]$
$\mathbf{k}$	Vector representation of a kernel tensor (linear filter)
K	Gain term applied to the exponents of CDI$^{\text{s}}$
K^{trans}	Volume transfer constant, a pharmacokinetic parameter derived from dynamic contrast-enhanced magnetic resonance imaging
$\mathcal{L}_{\mathbf{soft}}(\mathcal{D}_n, \mathcal{D}_p; \beta)$	Soft-AUC loss function with parameter β
$M(\mathbf{x}; \mathbf{b}, \boldsymbol{\rho})$	Weighted product of multiple diffusion-weighted intensities at location $\mathbf{x}$
$M_s(\mathbf{b}, \boldsymbol{\rho}; \boldsymbol{\theta})$	Weighted product of multiple synthetic diffusion-weighted intensities
$\mathbf{s}(\cdot)$	Softmax function
$S(b; \boldsymbol{\theta})$	General signal decay function in diffusion-weighted imaging
$S_m(b; S_0, A)$	Monoexponential model function in diffusion-weighted imaging
$S_s(b; S_0, D, \alpha)$	Stretched exponential model function in diffusion-weighted imaging
$S_0, S_0(\mathbf{x})$	Signal intensity with no diffusion weighting ($b = 0$ s/mm^2) in diffusion-weighted imaging
$\hat{S}_0, \hat{S}_0(\mathbf{x})$	Estimated signal intensity with no diffusion weighting ($b = 0$ s/mm^2) in diffusion-weighted imaging
$S_b, S_b(\mathbf{x})$	Signal intensity with b-value b in diffusion-weighted imaging
$\hat{S}_b, \hat{S}_b(\mathbf{x})$	Estimated signal intensity with b-value b in diffusion-weighted imaging
$\tilde{S}_b, \tilde{S}_b(\mathbf{x})$	Measured or estimated signal intensity with b-value b in diffusion-weighted imaging
$V(\mathbf{x})$	Local 3D volume centered at $\mathbf{x}$
$\mathbf{x}$	Location in 3D space

Chapter 1

Introduction

The prostate is a walnut-sized gland in the male reproductive system, located below the bladder and in front of the rectum [1]. The prostate is responsible for producing approximately one third of seminal fluid, an important component of semen. Over time, and especially with increasing age, the prostate may develop a number of diseases, including common benign conditions such as benign prostatic hyperplasia (BPH) and prostatitis as well as potentially life-threatening prostate cancer (PCa) [2].

PCa is the second most common form of cancer among men worldwide and the most frequently diagnosed cancer among men in in 112 countries, with roughly 1.4 million new cases in 2020 [3]. While the overall 5-year survival rate for prostate cancer is very high, prognosis is poor for patients with distant metastases outside of the prostate [4, 5]. As such, early diagnosis of PCa is critical for timely treatment of patients with PCa. However, many PCas develop slowly and pose a minimal risk of PCa-related mortality, in which case treatment can be limited to active surveillance of tumour development.

1.1 Prostate Cancer Screening and Diagnosis

Clinical screening for PCa has traditionally involved the use of prostate-specific antigen (PSA) screening, with high PSA levels used as an indicator of PCa [6]. Unfortunately, studies have shown that PSA-based screening has led to significant over-diagnosis of PCa, resulting in over-treatment that carries significant risks [7–9].

Diagnostic imaging has been increasingly used for PCa screening and diagnosis, with a variety of imaging modalities being employed. Transrectal ultrasound (TRUS) is routinely

used to guide prostate biopsies; however, its utility for PCa screening and diagnosis is limited due to the fact that PCa tumours are often isoechoic and thus cannot be delineated from surrounding tissue via TRUS. As a result, PCa screening and diagnosis using TRUS has low sensitivity and specificity [10]. PCa screening and diagnosis using positron emission tomography (PET) has also been explored, with several tracers showing promise for delineating cancerous and non-cancerous tissue in the prostate gland [11–14]. Unfortunately, the high cost of PET scanning makes it impractical as a diagnostic tool early in the screening pathway.

Magnetic resonance imaging (MRI) has grown significantly in prevalence for the purpose of PCa screening, with wide acceptance of the standardized Prostate Imaging–Reporting and Data System (PI-RADS) [15]. T2-weighted (T2w) MRI has been well-studied for PCa screening and diagnosis [16–19], where potentially cancerous regions are characterized by signal hypointensity, and is considered the primary determining modality for the transition zone (TZ) in the PI-RADS framework [15, 19]. However, T2w signal hypointensity in the peripheral zone (PZ) of the prostate gland can also be associated with a number of non-cancerous abnormal conditions such as inflammation, fibrosis, and hemorrhage [20, 21], leading to false positives if T2w is the sole method used. To improve diagnostic accuracy when using MRI for PCa assessment, two complementary MRI techniques have been leveraged alongside T2w: diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE) imaging [20]. These techniques, when used together with T2w, form a multiparametric magnetic resonance imaging (mpMRI) strategy to overcome the shortcomings of each modality. However, the need to interpret several modalities can increase interpretation challenges, resulting in increased inter- and intra-observer variability. Figure 1.1 shows example images for each mpMRI modality.

Recently, a new MRI technique called correlated diffusion imaging (CDI) [22] was proposed for improving PCa diagnosis. Preliminary studies demonstrated the potential of CDI for delineating between cancerous and non-cancerous tissue [22, 23]. However, the scopes of these studies are limited in terms of patient cohort size and diversity. Furthermore, a number of limitations exist in CDI as first introduced with respect to signal-to-noise ratio (SNR) and DWI acquisition time, as well as signal intensity (SI) variability in inter-patient and intra-patient acquisitions.

To accurately diagnose PCa, the gold standard technique is a prostate biopsy, where samples of tissue are removed from the prostate and examined under a microscope [24]. Typically, each tissue sample is assigned a Gleason score (GS) which reflects the cellular patterns present in the sample. To do so, the dominant and second-most-dominant cellular patterns in the sample are each assigned a grade from 1 to 5, with each grade representing a different cellular pattern and the overall scale representing a spectrum from healthy cells

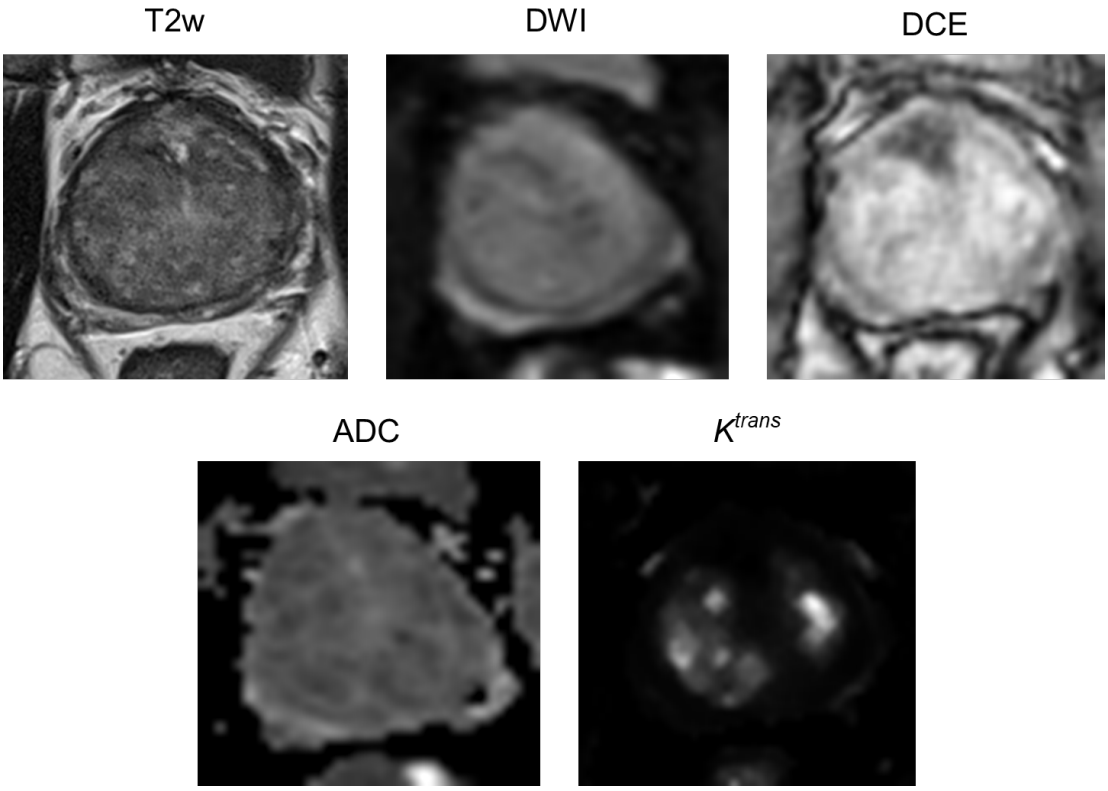

Figure 1.1: T2w, DW, and DCE images of the prostate (first row), as well as quantitative parameter maps derived from them (second row). ADC is calculated from DW images, while K^{trans} is calculated from DCE images.

to tumour cells [24]. Adding the two grades yields the final GS, which loosely reflects the risk the tumour presents [24]. While biopsies are a valuable tool for diagnosis and characterization of PCa, they may also cause discomfort and can cause serious infections or urine retention to occur in up to 1% of cases [7]. As such, non-invasive screening methods are preferable when possible, and reducing the number of biopsies through image-based guidance can also help minimize these risks.

1.2 Clinical Significance of Prostate Cancer

As alluded to in the previous section, PCa can vary considerably in the level of risk it presents, and not all PCas are considered to be life-threatening. Indeed, many low-grade

PCas may be considered clinically insignificant, meaning that the risk of PCa-related mortality is so low that it is not in the patient's best interest to actively treat the PCa. For example, it may be that PCa-related mortality is expected to occur beyond the patient's life expectancy, and in such cases treating the PCa could cause unnecessary stress, discomfort, pain, or side effects [7, 25]. This has given rise to a distinction between clinically insignificant prostate cancer (insPCa) and clinically significant prostate cancer (csPCa) based on the level of risk a tumour presents.

To better characterize risk levels, the International Society of Urological Pathology (ISUP) proposed the use of a series of grade groups (GGs) in 2014 [26], which was adopted by the World Health Organization (WHO) in 2016 [27] and reaffirmed by ISUP in 2019 [28]. The ISUP GG system divides specific Gleason patterns into 5 groups with progressively increasing risk levels [29]. This system better reflects prognosis; for example, GS 7 may represent patterns of 3+4 (GG 2) or 4+3 (GG 3), but the latter pattern exhibits significantly worse prognosis [26]. Building upon this, the National Comprehensive Cancer Network (NCCN) defined a risk stratification system which considers both GS and GG alongside additional information such as clinical stage or PSA level [29]. Table 1.1 provides additional detail on the correspondences between GS, GG, and risk group.

Table 1.1: Correspondences between NCCN risk groups, ISUP grade groups, and Gleason scores in the assessment of prostate cancer. Note that risk groups are not fully defined by Gleason scores or grade groups [29].

Risk Group	**Grade Group**	**Gleason Score**
Very low/low	1	≤ 6
Intermediate	2	7 (3+4)
	3	7 (4+3)
High/very high	4	8
	5	9-10

Although the conceptual notions of insPCa and csPCa are generally understood, there is disagreement and inconsistency in how the difference is clinically defined and identified [30]. The most common criteria used to define insPCa include the absence of Gleason patterns 4 or 5 (i.e., GG 1), organ-confined disease (i.e., PCa which has not spread beyond the prostate), and tumour volume less than some threshold (e.g., 0.5 mm^3 [25]) [30]. Notably, these criteria are not necessarily used in conjunction; GG 1 tumours are generally

considered insPCa, with treatment often deferred in favour of active surveillance [29, 31].

An important distinction, as noted by Mertens et al. [30], is that the designation "insPCa" does not imply that a tumour is irrelevant; indeed, active surveillance is required at a minimum, and some insPCa tumours may eventually be reclassified (in up to $\sim$1/3 of cases [31, 32]) or require treatment. However, past studies of active surveillance have noted that grade reclassification was less likely for patients who underwent mpMRI [31], and that mpMRI findings may be helpful for selecting and monitoring patients in active surveillance programs [33].

A notable challenge in PCa treatment planning is that many patients diagnosed with GG 1 PCa may actually have higher-grade PCa; biopsy-based grading of PCa is sensitive to the particular biopsies taken, meaning that higher-grade PCa may not be sampled [9]. Indeed, tumours which undergo early reclassification during active surveillance more likely represent sampling errors as opposed to significant changes in tumour characteristics [32]. MRI presents an opportunity to identify such cases, as magnetic resonance (MR) images provide dense measurements within the prostate [9]. However, this is predicated on the ability of MRI to effectively differentiate between low-grade and high-grade cancers.

1.3 Thesis Contributions

In an effort to improve MRI-based PCa screening and risk assessment, the two main contributions of this thesis are:

1. A DWI-based computational MRI technique called synthetic correlated diffusion imaging (CDI^s), which extends CDI through the use of synthetic DWI. CDI^s multiplicatively combines both native and synthetic DWI acquisitions in a weighted fashion, allowing a user to tailor the particular weights for specific applications. To improve inter- and intra-patient consistency, a per-patient calibration factor is also introduced.

2. A gradient-based optimization framework through which the parameters involved in the calculation of CDI^s may be automatically optimized to improve performance. In this thesis, the framework is used to optimize CDI^s for PCa screening and risk assessment applications, but in general any appropriate objective function may be optimized using the proposed framework.

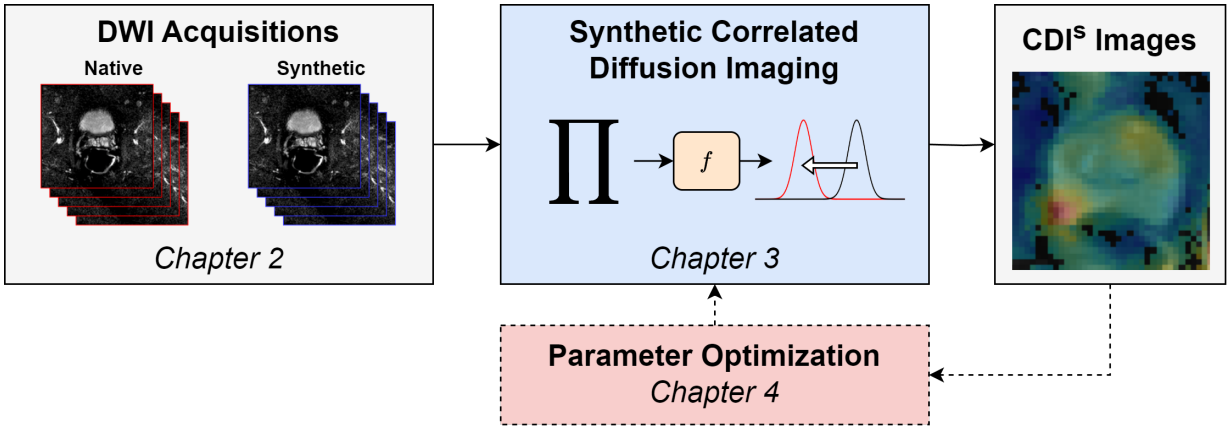

Figure 1.2: The CDI^s optimization framework proposed in this thesis. Native and synthetic DWI images are used to compute a CDI^s image. Through iterative optimization, the parameters of CDI^s are tailored to produce better images for downstream clinical tasks.

Figure 1.2 illustrates how the chapters of this thesis are connected. First, Chapter 2 discusses the principles and applications of DWI, synthetic DWI, and CDI in PCa imaging, including relevant related works. Chapters 3 and 4 then present the main contributions of this thesis, including their development and performance in experiments. Lastly, a summary of the thesis and limitations of the proposed methods are presented in Chapter 5, as well as potential directions for future research.

Chapter 2

Background

This chapter discusses the relevant background in physics, biology, and MRI for this thesis, as well as previous studies related to the proposed thesis contributions. Section 2.1 discusses DWI, including its physical principle, relevant models of DW signal decay, and related works which explore the use of DWI for prostate cancer diagnosis and grading. Building upon this, Section 2.2 introduces the notion of synthetic DWI and explains how it can be computed, as well as how it has previously been applied to PCa. Section 2.3 explains the theory and definition of CDI, a computational MRI modality derived from DWI. Previous studies of CDI are also discussed in this section.

2.1 Diffusion-weighted Imaging

DWI is a form of MRI which measures the Brownian motion (i.e., diffusion) of water molecules in biological tissue. Diffusion parameters estimated from DWI are routinely used in clinical practice for diagnosis and assessment of various diseases, such as cancer [34, 35], fibrosis [36], and stroke [37, 38]. In this section, the physical principle, measurement techniques, and applications of DWI are briefly described.

2.1.1 Physical Principle

Liquid water molecules experience random Brownian motion driven by thermal energy, causing the molecules to diffuse randomly. When barriers are introduced (e.g., membranes

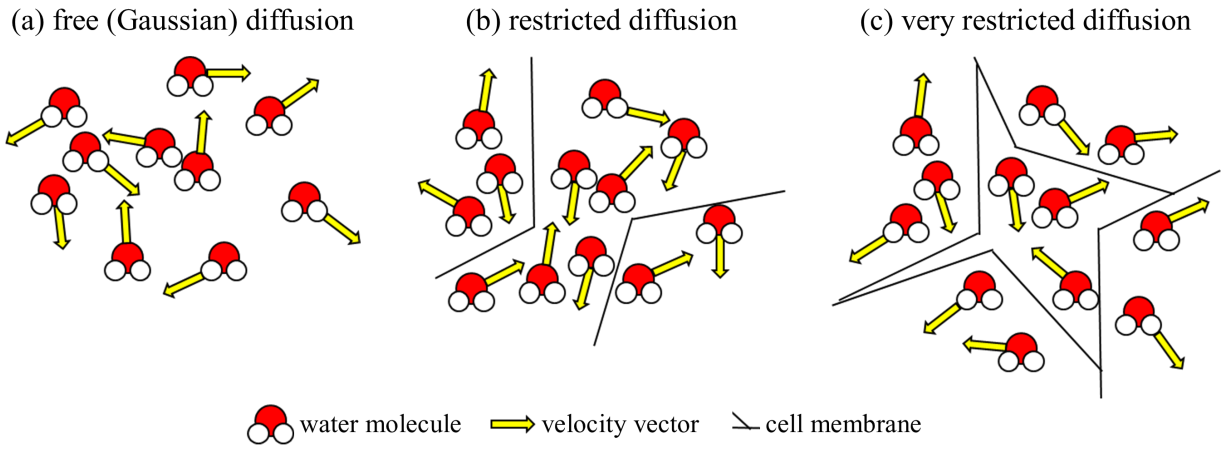

Figure 2.1: Conceptual illustration of restricted diffusion in biological tissues. (a) In the absence of barriers, water diffuses randomly in all directions. (b) In the presence of cell membranes, water diffusion across the membranes is reduced. (c) As the density of cell membranes increases, diffusion is further restricted; the degree of diffusion restriction reflects the cellular density and architecture.

or walls), water molecules are no longer free to diffuse randomly, and consequently the diffusion behaviour of water molecules is largely dictated by their surroundings.

Water makes up a large proportion of biological tissues, both within cells (intracellular fluid) and between cells (extracellular fluid). In human tissues, water diffusion is primarily restricted by cell membranes, and as a result the cellular architecture and cellular density strongly influence the diffusion characteristics of tissues [39]. For example, the rapid cellular growth exhibited by cancerous tissues causes cells to be tightly-packed, which restricts water diffusion. This effect is illustrated in Figure 2.1

Like all elementary particles, the protons in water molecules (i.e., hydrogen atoms) exhibit a spin, which consequently produces a small magnetic dipole due to their positive charge [40]. In the absence of an external magnetic field, these dipoles are oriented randomly and have zero net magnetization. However, when acted on by an external magnetic field, the magnetic dipoles preferentially align with the applied field and produce a net magnetization, which then decays back to zero when the external field is removed [40].

In DWI, a pair of so-called diffusion-sensitizing magnetic gradients are used to produce differences in SI based on differences in diffusion characteristics. Generally, symmetric magnetic gradients are applied before and after a 180 degree pulse – the first gradient de-phases the spins of the protons by imbuing them with different phase shifts, while

the second re-phases them by effectively removing the phase shift [41]. However, moving water molecules (i.e., those undergoing diffusion) are not subject to precisely the same gradient in both directions, and as such they re-phase imperfectly and cause a reduction in net magnetization. The amount of phase misalignment reflects the amount of Brownian motion, and hence the MR signal is attenuated by increased diffusion [41]. Conversely, the resulting DW images exhibit greater SI in areas of greater diffusion restriction, thereby providing useful contrast.

Generally, diffusion-sensitizing gradients are applied in several different directions to measure differences in directional diffusion. While the diffusion directions may be considered (e.g., in diffusion tensor imaging), in many applications the overall diffusion is of interest. To produce isotropic DW images, signals acquired in each gradient direction are combined via their geometric mean [42].

2.1.2 Signal Decay Models

Although DWI produces diffusion-based tissue contrast, it is still a largely qualitative technique in its basic form. However, for better characterization of tissues, the physical diffusion parameters are usually of interest.

The pulse sequences used in DWI are characterized by b-values which reflect the strengths and timings of the gradient pulses [41]. A b-value may be expressed as

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right), \tag{2.1}$$

where γ denotes the proton gyromagnetic ratio, G represents the gradient pulse strength, δ represents the gradient pulse duration, and Δ represents the time between gradient pulses [41, 43]. DW SI decays as a function of b, and so several DW acquisitions at different b-values may be used to characterize the decay behaviour.

While the exact signal decay behaviour of DWI as a function of b is challenging to characterize, it may be approximated by simpler decay models. By fitting a decay model to measured signal intensities at various b-values, quantitative tissue parameters may be estimated for use in diagnosis and grading of prostate cancers. In general, signal decay models may be represented by $S(b; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents the parameters of the decay model.

In this section, two commonly-used models of DW signal decay are introduced. While several other models of DW signal decay have been explored in literature [44–47], these

models are more complex and typically require acquisitions at high b-values (e.g., up to 2000-3000 s/mm^2) or at many different b-values to be accurately fit to measured signals.

Monoexponential Model

The simplest model of DW signal decay is the monoexponential model (MEM), which assumes Gaussian diffusion behaviour (i.e., free diffusion). Under the MEM, SI decays according to

$$S_m(b; S_0, A) = S_0 \exp(-bA) \tag{2.2}$$

where $S_m(b; S_0, A)$ denotes SI at b-value b, S_0 denotes the SI at $b = 0$ s/mm^2, and A denotes the ADC [34]. Equation (2.2) is log-linear, allowing for S_0 and A to be estimated via linear regression in log space. Notably, the MEM can be considered either a 2-parameter or 1-parameter model depending on whether S_0 is considered to be a measurement or a parameter. The term "apparent diffusion" refers to the fact that biological tissues do not exhibit true Gaussian diffusion, and ADC may vary depending on the b-values used to estimate it.

ADC is routinely used in clinical practice for both diagnosis and characterization of PCa, and is one of the core modalities included in the PI-RADS framework [15]. ADC values may be used to differentiate PCa tissue from normal tissue or benign prostate conditions [48, 49], and are correlated with GS [49]. Moreover, ADC may aid in assessment of tumour aggressiveness, which can help to distinguish between insPCa and csPCa and identify patients eligible for active surveillance [50, 51].

Stretched Exponential Model

The stretched exponential model (SEM) adds a parameter $\alpha \in [0, 1]$ to the monoexponential decay model to account for non-log-linear behaviour. Under the SEM, SI decays according to

$$S_s(b; S_0, D, \alpha) = S_0 \exp(-(bD)^\alpha) \tag{2.3}$$

where $S_s(b; S_0, D, \alpha)$ denotes SI at b-value b, S_0 denotes the SI at $b = 0$ s/mm^2, and D denotes the distributed diffusion coefficient (DDC), which represents mean intravoxel diffusion rates [52, 53]. $\alpha \in [0, 1]$ represents the water molecular diffusion heterogeneity [52, 53]. Notably, the SEM can be considered either a 2-parameter or 3-parameter model depending on whether S_0 is considered to be a measurement or a parameter. In DWI literature, standard iterative non-linear optimization algorithms are used almost exclusively for SEM fitting. Of these algorithms, the Levenberg-Marquardt (LM) algorithm [54, 55]

is by far the most common [49, 56–61], however Gauss-Newton [52] and quasi-Newton [62] approaches have been explored as well.

The SEM for DW signal decay was first proposed by Bennett et al. [52], and has since been used to model DW signal decay in a variety of human tissues. In PCa applications, the SEM may better characterize DWI signal decay [59] and has been explored for PCa detection [48, 58, 63], PCa grading [49], and evaluation of csPCa [53]. Although SEM parameters exhibit correlation with GS and tissue composition parameters [49, 63], many studies have found that the SEM may not have clinical value beyond that provided by the MEM [48, 49, 53, 58].

2.2 Synthetic Diffusion-weighted Imaging

DW acquisitions are limited to the specific b-values acquired during the MRI study, with protocols varying from institution to institution. Although more b-values can be acquired, scans are typically limited to the most diagnostically useful b-values for a particular anatomical structure or pathology in order to minimize scanning time. Moreover, higher b-values take longer to acquire (introducing increased risk of distortion due to patient motion) and typically exhibit lower SNR, which can limit the usability of these scans in practice [64, 65].

To allow for additional b-values to be used without being specifically acquired, it is possible to use a decay model to extrapolate from real DWI acquisitions, producing synthetic DWI acquisitions. In clinical practice, synthetic acquisitions are often seen in the form of computed high b-value diffusion-weighted imaging (CHB-DWI), where low b-value acquisitions are used to estimate a high b-value acquisition. This allows for reduced acquisition time and better image quality, while retaining the tissue contrast characteristics of high b-value acquisitions [64, 65].

Figure 2.2 illustrates the general process by which synthetic DWI acquisitions are obtained. For the MEM, synthetic acquisitions are obtained by plugging estimates $\hat{S}_0$ and $\hat{A}$ into the MEM function defined in Equation (2.2). Specifically,

$$S_m(b; \hat{S}_0, \hat{A}) = \hat{S}_0 \exp(-b\hat{A}) \tag{2.4}$$

Notably, $\hat{S}_0$ may be obtained in different ways. Most intuitively, it may be measured directly as the SI at $b = 0$ s/mm^2 (i.e., S_0, in which case $\hat{S}_0$ is a measurement rather than an estimate). Additionally, during fitting of Equation (2.2), $\hat{S}_0$ may be obtained from the intercept of the log-linear fit. Given a reference acquisition S_{b_r} at b-value b_r, $\hat{S}_0$ may also

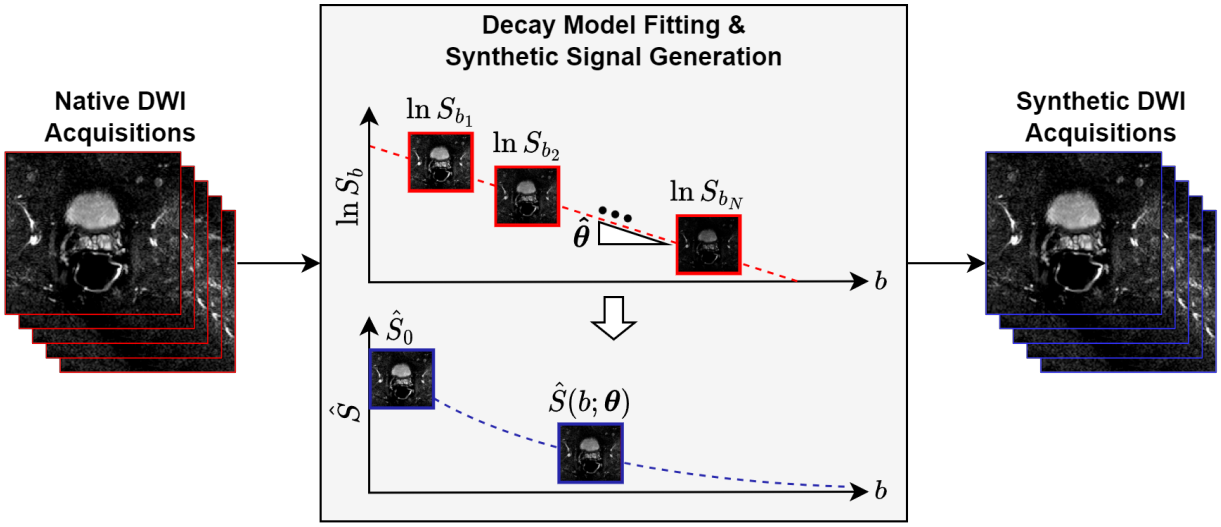

Figure 2.2: Generation of synthetic DWI acquisitions. A signal decay model is fit to a series of native DWI acquisitions at different b-values, voxel-wise. Once fitting is performed, any b-value can be sampled along the voxel-wise decay curves to produce a synthetic DWI acquisition.

be obtained by rearrangement of Equation (2.2):

$$\hat{S}_0 = S_{b_r} \exp(b_r \hat{A}) \tag{2.5}$$

For the SEM, synthetic acquisitions are obtained by plugging estimates $\hat{S}_0$, $\hat{D}$, and $\hat{\alpha}$ into the SEM function defined in Equation (2.3). Specifically,

$$S_s(b; \hat{S}_0, \hat{D}, \hat{\alpha}) = \hat{S}_0 \exp(-(b\hat{D})^{\hat{\alpha}}) \tag{2.6}$$

As in the MEM case, $\hat{S}_0$ may be obtained in a few ways. It may be measured directly as a DWI acquisition, or estimated when S_0 is treated as a parameter during fitting of Equation (2.3). From a reference acquisition S_{b_r} at b-value b_r, $\hat{S}_0$ may also be estimated by rearrangement of Equation (2.3):

$$\hat{S}_0 = S_{b_r} \exp((b_r \hat{D})^{\hat{\alpha}}) \tag{2.7}$$

CHB-DWI using the MEM has been widely explored in DWI literature. Ning et al. [66] found that CHB-DWI images at $b = 2000$ s/mm^2 can provide increased PCa sensitivity compared to native acquisitions at $b = 1200$ s/mm^2, although at the expense of reduced

specificity. Similarly, Rosenkrantz et al. [67] found that CHB-DWI can improve image quality and tumour detection over native DWI acquisitions. A later study by Rosenkrantz et al. [68] examined CHB-DWI with b-values up to 5000 s/mm^2 and found that the optimal b-values for PCa detection were in the range of 1500-2500 s/mm^2. A similar range was echoed by Agarwal et al. [69], who found that $b = 1600$ s/mm^2 was optimal for identifying patients eligible for active surveillance (i.e., insPCa patients), and that b-values in the range of 1067-2000 s/mm^2 exhibited no significant differences in performance.

2.3 Correlated Diffusion Imaging

First proposed by Wong et al. [22], CDI is a computed MRI modality based on DWI where the tissue being imaged is characterized by the joint correlation of DW signal signal attenuation across multiple gradient pulse strengths and timings (i.e., b-values). By incorporating DW signals with sensitized to different degrees of water diffusion, CDI may better discriminate PCa tissue and healthy tissue compared to standard MRI modalities.

In CDI, a series of DWI acquisitions with different b-values are used to probe water molecules with different degrees of Brownian motion within a local volume. Signal mixing is then performed to determine the joint correlation of the acquisitions within a local volume. As such, CDI leverages the distribution of water molecules with different degrees of Brownian motion within the local volume to delineate between cancerous tissue (indicated by signal hyperintensity) and non-cancerous tissue (indicated by signal hypointensity).

CDI is defined by a product of acquired DWI signal intensities modified by a prior probability density function (PDF). Given acquired b-values $\mathbf{b} = \begin{bmatrix} b_1 & \cdots & b_N \end{bmatrix}^\top$, CDI SI is given by

$$C(\mathbf{x}; \mathbf{b}) = \iiint_{V(\mathbf{x})} \prod_{i=1}^{N} \Big(S_{b_i}(\mathbf{x}') \Big) f(\mathbf{x}') d\mathbf{x}' \tag{2.8}$$

where $\mathbf{x}$ denotes spatial location, S_{b_i} denotes an acquired signal at b-value b_i, $V(\mathbf{x})$ represents a local volume around $\mathbf{x}$, and f denotes a PDF over $V(\mathbf{x})$ which may be dependent on spatial location (i.e., $\mathbf{x}'$ and $\mathbf{x}$) and DWI SI [22].

Wong et al. [22] previously compared CDI with ADC for delineation of healthy and cancerous tissues, finding that CDI achieved a marked increase in area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy [22]. This study also visually compared T2w, ADC, and CDI, finding that CDI produced clearer visual indications of PCa compared to the other modalities based on radiologist assessment.

However, this study was limited to 20 subjects and did not examine differences between insPCa and csPCa tissues.

Khalvati et al. [23] extracted texture features from CDI, DWI, ADC, CHB-DWI, and T2w images and used these features to train support vector machines for classification of PCa. It was found that CDI features alone outperformed multi-modal features from mpMRI (i.e., T2w + ADC) and mpMRI + CHB-DWI, and incorporating CDI into multi-modal feature models markedly improved classification performance. However, this study only considered tumours with GS ≥ 7 (i.e., insPCa was not considered), and was limited to a cohort of 20 subjects.

2.4 Summary

This chapter provided the necessary background knowledge to understand how and why CDI is extended to CDI^s in the next chapter. Specifically, the fundamental concepts of DWI were introduced, including its physical principles, common signal decay modelling techniques, and extrapolation of DWI decay models to produce synthetic DW images. The clinical applications of DWI and DWI-derived diffusion parameters were also discussed to better understand its value in clinical PCa assessment. Furthermore, the theory and calculation of CDI, a DWI-based computational modality, were explained. Previous studies comparing CDI to existing MRI modalities were presented, illustrating its potential as a PCa assessment tool and highlighting the gaps in CDI research which this thesis will aim to address.

Chapter 3

Synthetic Correlated Diffusion Imaging

In this chapter, an extension of CDI called CDI^s is proposed to address the limitations of CDI, which are discussed in Section 3.1. The key steps of CDI^s are weighted signal mixing, application of a prior PDF, and signal calibration, which are described in Section 3.2, Section 3.3, and Section 3.4, respectively. After incorporating these steps, the final form of CDI^s is defined in Section 3.5. The patient cohort and MRI data used in this thesis are described in Section 3.6. Section 3.7 describes the experimental setup and applicable quantitative and qualitative metrics used to analyse CDI^s's potential clinical utility, with experimental results presented in Section 3.8. A brief summary is provided in Section 3.9.

3.1 Problem Formulation

While preliminary studies have shown that CDI holds considerable promise of achieving greater signal delineation between cancerous and non-cancerous tissue when used as a standalone diagnostic imaging method [22] and when used in combination with T2w and DWI [23], several challenges remain which limit the utility of CDI in practice.

CDI's flexibility is limited by the b-values of the DWI acquisitions used to compute it. Changing the particular b-values or adding more b-values corresponds to a change in imaging protocol, which may be challenging for institutions with established DWI protocols for PCa-related investigations. This means that CDI may not perform optimally (in terms of diagnostic or prognostic value) depending on the particular b-values used. Furthermore,

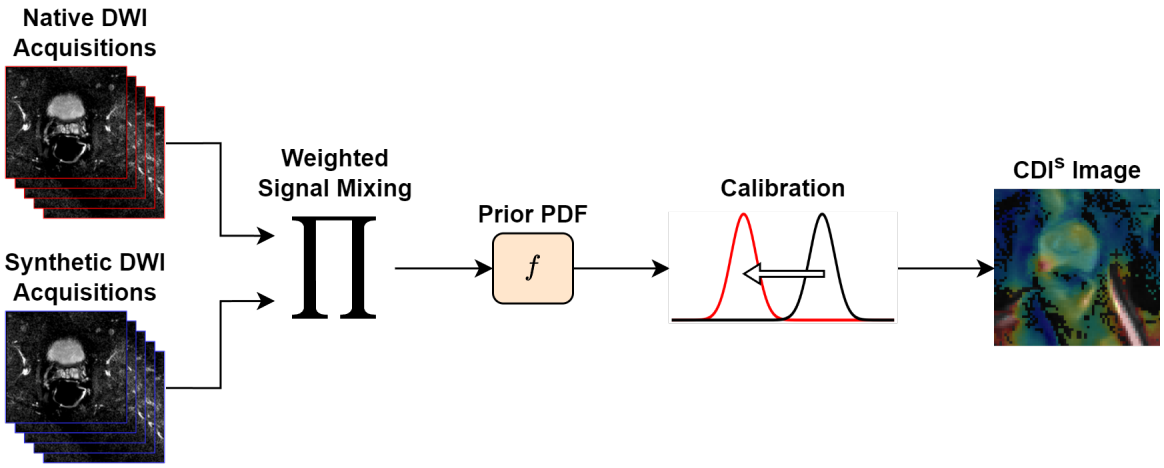

Figure 3.1: The proposed formulation of CDI^s. Native and synthetic DWI acquisitions at different b-values are weighted and multiplied, then filtered using a prior PDF over the neighbourhood of each voxel. The resulting image is then calibrated to produce a final CDI^s image.

when different protocols are used (e.g., when subjects in a cohort were scanned at multiple institutions), the inflexibility in b-values means that CDI SI cannot be meaningfully compared between the differing protocols.

Even without large differences in imaging protocol, CDI SI may still suffer from inter- and intra-patient variability due to differences in MRI scanner hardware. Different MRI scanners may acquire DW images differently (e.g., using different pulse sequences, field strengths, or coils), and due to the arbitrary units of DWI SI the scans are not generally comparable in their original form.

Past studies of CDI [22, 23] have been limited in scope, as the cohorts examined were relatively small (both studies examined 20 subjects). Moreover, these studies have not directly compared CDI to MRI modalities commonly used in clinical practice, such as T2w and DCE (including DCE-derived K^{trans} maps), and the use of CDI for differentiating between csPCa and insPCa has not been explored.

To address the aforementioned limitations of CDI, an extended form of CDI called CDI^s is proposed. CDI^s uses synthetic signal acquisitions alongside native DWI acquisitions to allow for flexibility in the choice of b-values, and also incorporates a mechanism to increase or decrease the influence of acquisitions at particular b-values. Furthermore, different priors (filters) are briefly introduced, and a new method of calibration is proposed to increase

the consistency of CDI^s SI. Figure 3.1 illustrates the method developed in this chapter at a high level, and the following sections discuss each component in more detail.

3.2 Weighted Signal Mixing

In CDI^s, synthetic DWI acquisitions are obtained via extrapolation of a DW signal decay model. These acquisitions may then be included in the signal mixing process as if they were native DWI acquisitions – in essence, the same basic signal mixing process may be used in CDI and CDI^s, but CDI^s incorporates both native and synthetic signals.

The signal mixing approach of CDI weighs each b-value according to its inherent SI, effectively performing a weighted product with uniform weights of 1. However, previous studies have shown that not all b-values are equally important for diagnosis and grading of PCa [68, 69]. Considering this, a mechanism is proposed by which b-values may be weighted differently to reflect their clinical value. Specifically, since CDI uses a uniformly-weighted product, a nonuniformly-weighted product may be used to boost signals from clinically-important b-values and attenuate those of less-important b-values. To achieve this, weighting exponents $\boldsymbol{\rho} = \begin{bmatrix}\rho_1 & \cdots & \rho_N\end{bmatrix}^\top$ are introduced and each exponent ρ_i is applied to its corresponding DWI acquisition:

$$M(\mathbf{x}; \mathbf{b}, \boldsymbol{\rho}) = \prod_{i=1}^{N} \tilde{S}_{b_i}(\mathbf{x})^{\rho_i} \tag{3.1}$$

In Equation (3.1), $\tilde{S}_{b_i}(\mathbf{x})$ denotes either a native DWI signal ($S_{b_i}(\mathbf{x})$) or synthetic DWI signal ($\hat{S}_{b_i}(\mathbf{x}) = S(b_i; \hat{\boldsymbol{\theta}}(\mathbf{x}))$) at b-value b_i.

3.3 Probability Density Functions

The function f in Equation (2.8) represents a PDF applied to the product of DWI acquisitions, which is analogously applied to $M(\mathbf{x})$ in CDI^s. f is defined as a prior probability distribution over $V(\mathbf{x})$, and it helps to impose structure and filter noise in the DWI acquisitions which has been amplified by the product in M. Formally, f may be any distribution over the region $V(\mathbf{x})$ and may include dependence on DWI SI.

One of the simplest forms of f is a Gaussian distribution centered at $\mathbf{x}$, defined as

$$f_g(\mathbf{x}'; \mathbf{x}, \Sigma) = \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}' - \mathbf{x})\Sigma^{-1}(\mathbf{x}' - \mathbf{x})^\top\right) \tag{3.2}$$

where $\mathbf{x}'$ is a location within $V(\mathbf{x})$, $\mathbf{x}$ is the center of the distribution, and Σ is a 3×3 covariance matrix. The use of a Gaussian means that locations in $V(\mathbf{x})$ distant from $\mathbf{x}$ have less influence on the SI at $\mathbf{x}$, thereby encoding a heuristic of local homogeneity.

A notable drawback of the Gaussian is that it is not edge-preserving, meaning that structural information is lost in the resulting CDI^s images. To alleviate this concern, an edge-preserving adaptive function may be used instead, with one such function being the bilateral filter. In this case, a function f_b may be written as

$$f_b(\mathbf{x}'; \mathbf{x}, \Sigma, \sigma_I) = \frac{1}{W(\mathbf{x})} f_g\left(\mathbf{x}'; \mathbf{x}, \Sigma\right) g\left(\ln\left(\frac{M(\mathbf{x}')}{M(\mathbf{x})}\right); 0, \sigma_I\right) \tag{3.3}$$

where g is a one-dimensional Gaussian distribution with mean 0 and standard deviation σ_I, f_g is as defined in Equation (3.2), and $W(\mathbf{x})$ is a normalizing term to ensure that f_b is a valid distribution over $V(\mathbf{x})$:

$$W(\mathbf{x}) = \iiint_{V(\mathbf{x})} f\left(\mathbf{x}'; \mathbf{x}, \Sigma\right) g\left(\ln\left(\frac{M(\mathbf{x}')}{M(\mathbf{x})}\right); 0, \sigma_I\right) d\mathbf{x}' \tag{3.4}$$

Intuitively, f_g acts as a spatial weighting function and g adjusts the weighting such that voxels with large intensity differences relative to $M(\mathbf{x})$ receive a lower weight. In this sense, f_g may be thought of as both a spatial and intensity prior.

3.4 Signal Calibration

In its original form, CDI does not include any form of calibration, meaning that the dynamic range of CDI may vary considerably between institutions, scanners, subjects, and even multiple scans of the same subject. As a result, it is difficult to compare CDI SIs directly unless imaging protocols and hardware are strictly enforced across all subjects, which is unrealistic in multi-institutional studies or large, diverse patient cohorts.

To address this limitation, a calibration factor is proposed for CDI^s which corrects for variability in DWI SI due to differences in the scanner and imaging protocol used during acquisition. The calibration factor is defined as the median SI after signal mixing and filtering, and the uncalibrated CDI^s values are divided by this factor. Examining Figure 3.2, the distribution of CDI SIs is observed to be approximately log-normal within the prostate. Considering this, dividing the CDI^s SI by its median corresponds to subtraction of the median in log space, which centers the distribution at approximately 0 in log space.

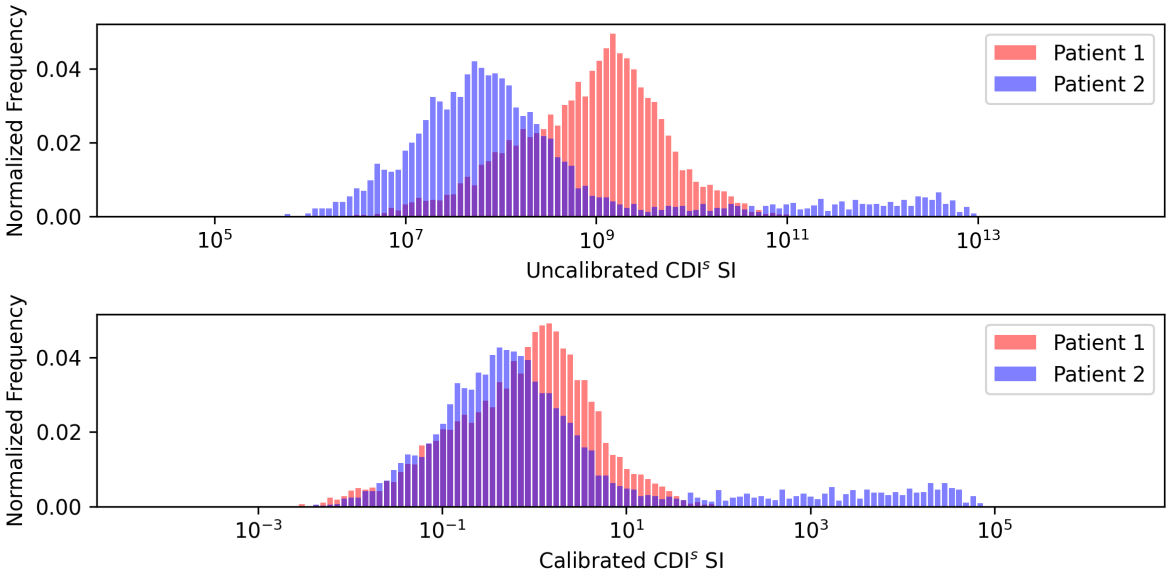

Figure 3.2: Histograms of within-prostate CDI^s SIs before and after calibration for two patient cases. Before calibration, the distributions have different means and very high SIs. After calibration, the distributions are both centered near 0 and have lower SIs.

The rationale for this calibration factor is that differences in DWI SI are primarily differences of global scale. This follows from the fact that $S_b \propto S_0$ in general (e.g., in Equation (2.2) and Equation (2.3)). As such, if S_0 is scaled by some constant $a > 0$ (as might happen when changing the MRI scanner or subject), S_b is scaled by the same factor while the shapes of the signal decay curves remain relatively unchanged. By dividing CDI^s SI by its median, any differences in global scale between different MRI studies are effectively removed (see Appendix A.1).

Since the objective is to calibrate signals within the prostate, it does not make sense to compute the true global median SI – indeed, anatomical structures in prostate scans such as the bladder and surrounding muscle have CDI^s SI distributions which differ significantly from that of the prostate. Ideally, the median would be computed from voxels within the prostate, but this would require time-consuming segmentation of the prostate. To avoid this, a cylindrical mask is defined with its central axis perpendicular to the slice planes of the DWI acquisitions and passing through the center of all slices, as illustrated in Figure 3.3. This mask allows for the central region of the CDI^s volume to be extracted and used for calculation of the median, thereby including the prostate and minimizing spurious contributions from other anatomical structures.

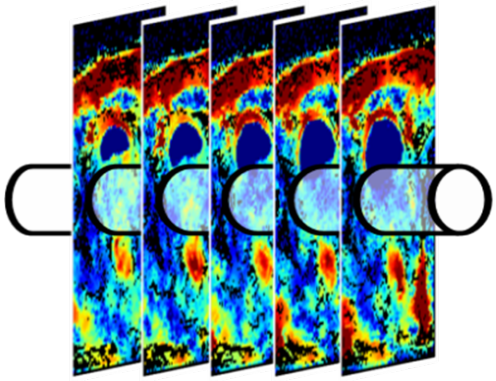

Figure 3.3: Illustration of the cylindrical mask used to compute the median CDI^s SI within the prostate for calibration.

3.5 Definition

By combining weighted signal mixing, a prior PDF, and calibration, CDI^s SI is given by

$$C_s(\mathbf{x};\mathbf{b},\boldsymbol{\rho}) = \frac{1}{Z}\iiint_{V(\mathbf{x})} M(\mathbf{x}';\mathbf{b},\boldsymbol{\rho}) f(\mathbf{x}') d\mathbf{x}' \tag{3.5}$$

where $\mathbf{x}$ represents spatial location, M is as defined in Equation (3.1), $\mathbf{b}$ and $\boldsymbol{\rho}$ represent b-values and corresponding coefficients for controlling their contributions (respectively), $V(\mathbf{x})$ represents a local volume around $\mathbf{x}$, f denotes a PDF over $V(\mathbf{x})$ (e.g., Equations (3.2) and (3.3)), and Z represents the calibration factor. In reality, DWI acquisitions consist of discrete voxels, and therefore in practice CDI^s is computed as

$$C_s(\mathbf{x};\mathbf{b},\boldsymbol{\rho}) = \frac{1}{Z}\sum_{\mathbf{x}'\in V(\mathbf{x})} M(\mathbf{x}';\mathbf{b},\boldsymbol{\rho}) f(\mathbf{x}') \tag{3.6}$$

where f is a discrete distribution over $V(\mathbf{x})$.

3.6 Patient Data

To study the relationship between CDI^s SI and PCa, a cohort of 200 patient cases with histopathology validation acquired at Radboud University Medical Centre (Radboudumc)

Table 3.1: Summary of age, MR scanner, and clinical significance variables of the patient cohort examined in this thesis. Age and MR scanner statistics are presented at the patient level, while clinical significance statistics are presented at the tumour level.

Age	
30-39	3 (1.5%)
40-49	5 (2.5%)
50-59	45 (22.5%)
60-69	112 (56%)
70-79	35 (17.5%)
MR Scanner (Siemens MAGNETOM)	
Skyra 3.0T	195 (97.5%)
Trio 3.0T	5 (2.5%)
Clinical significance (Gleason Score)	
csPCa (GS $\geq$ 7)	76 (25.4%)
insPCa (GS $<$ 7 or PI-RADS $\leq$ 2)	223 (74.6%)

in the Prostate MRI Reference Center in Nijmegen, The Netherlands [70–72] were used in this thesis. Notably, findings with a PI-RADS score of 1 or 2 were not biopsied and were considered clinically insignificant. Table 3.1 summarizes the demographic, MR scanner, and clinical significance variables of the patient cohort used in this thesis. The patients in this cohort ranged in age from 37-78 years, with a median age of 64 years. All acquisitions were performed using a Siemens MAGNETOM Trio 3.0T machine or a Siemens MAGNETOM Skyra 3.0T machine, and were reviewed by or performed under the supervision of an expert radiologist with over 20 years of experience interpreting prostate MRI [70].

A single-shot echo-planar sequence was used for axial DWI acquisitions, with TR ranging from 2500-3300 ms with a median of 2700 ms and TE ranging from 63-81 ms with a median of 63 ms. The in-plane resolution of the acquisitions was 2 mm with slice thickness ranging from 3-4.5 mm with a median of 3 mm. The display field of view (DFOV) ranged from 16.8×25.6 cm^2 to 24.0×25.6 cm^2 with a median of 16.8×25.6 cm^2, and images were acquired at b-values of 50 s/mm^2, 400 s/mm^2, and 800 s/mm^2. To compare the performance of CDI^s for PCa delineation with standard MRI techniques, ADC maps were also obtained from DWI acquisitions.

Axial T2w acquisitions were also obtained as a reference of comparison, and were

performed using a turbo spin-echo sequence with TR ranging from 3880-7434.8 ms with a median of 5660 ms and TE ranging from 101-112 ms with a median of 104 ms. The in-plane resolution of the acquisitions ranged from 0.3-0.6 mm with a median of 0.5 mm and slice thickness ranged from 3-4.5 mm with a median of 3 mm. The DFOV ranged from 18×18 cm^2 to 19.2×19.2 cm^2 with a median of 19.2×19.2 cm^2.

Axial DCE imaging was conducted with a turbo flash gradient-echo sequence, with TR ranging from 3.72-36 ms with a median of 36 ms and TE ranging from 1.41-1.84 ms with a median of 1.41 ms. The in-plane resolution of the acquisitions ranged from 1.3-1.8 mm with a median of 1.5 mm, slice thickness ranged from 3-5 mm with a median of 3.5 mm, and the temporal resolution was 3.5 s. The DFOV ranged from 19.2×19.2 cm^2 to 25×25 cm^2 with a median of 19.2×19.2 cm^2. Maps of the pharmacokinetic parameter K^{trans} were obtained from the DCE series.

PCa, whole gland, TZ, and PZ annotations for all patient acquisitions in this cohort were used in this thesis, with the annotation being performed by two radiology residents and two experienced board-certified radiologists (working in pairs) at the University of Naples Federico II, Naples, Italy [73].

3.7 Experimental Setup

CDI^s was computed from the DWI acquisitions described in the previous section using b-values $\mathbf{b} = \begin{bmatrix}0 & 50 & 400 & 800 & 1000 & 2000 & 3000 & 4000 & 5000\end{bmatrix}^\top$ and exponents $\boldsymbol{\rho} = \mathbf{1}_9$ (a 9-vector of ones). Both Gaussian and bilateral filters were explored; $\sigma_I = 4$ was used for the bilateral filter and $\Sigma = \text{diag}(4\text{ mm}^2, 4\text{ mm}^2, 0\text{ mm}^2)$ was used for both filters. The cylindrical calibration mask described in Section 3.4 was used in all experiments with a radius of $0.25\min(H, W)$, where H and W are the height and width of the DWI slices, respectively.

To investigate the relationship between CDI^s SI and the presence of PCa, a receiver operating characteristic (ROC) curve analysis was performed using CDI^s to quantitatively assess the ability to delineate between healthy tissue, csPCa tissue, and insPCa tissue. csPCa tissue is defined as tissue with a GS greater than or equal to 7 (GG 2-5) while insPCa tissue is defined as tissue with a GS less than 7 (GG 1) [25, 71]. The ROC curves were estimated empirically, and for illustrative purposes ROC curves obtained from the pooled data of all patient cases were plotted. Two clinical tasks are considered: identification of PCa tissue (i.e., separation of healthy tissue from insPCa and csPCa tissue, for PCa screening) and delineation of insPCa and csPCa tissue (for PCa treatment planning).

To provide a quantitative assessment, the area under the ROC curve was obtained as a single metric of delineation performance. For comparison purposes, curve analysis was also performed using T2w, DWI-derived ADC map values, and DCE-derived K^{trans} map values.

To assess the statistical significance of differences in AUC values between different modalities, the estimates of AUC variance and covariance proposed by DeLong et al. [74] are adopted, with two-tailed z-tests performed in all cases and 95% confidence intervals (CIs) constructed from the variance estimates. Notably, due to differences in image resolution, tests involving either T2w or K^{trans} could not be performed in a paired fashion and were therefore performed as unpaired tests. The unpaired tests produce more conservative p-values, meaning that there is an increased risk of type II error. However, if the null hypothesis is rejected, the use of unpaired tests presents no issue. An unadjusted significance level of 0.05 was initially selected; however, since 21 tests are being performed (i.e., tests between every possible pairing of modalities), the Bonferroni correction [75] was applied, yielding an adjusted significance level of $0.05/21 = 0.00\overline{238095}$.

Given that CDI^s is computed as a product of exponential signal acquisitions, CDI^s SI is more naturally interpreted in log space. As such, for visualization purposes during clinical interpretation, CDI^s SI is displayed in log space. Furthermore, the transformed CDI^s SI is visualized as a heatmap overlay on T2w images to provide additional anatomical context with respect to the prostate gland. DWI-derived ADC maps and DCE-derived K^{trans} maps are also visualized as heatmap overlays on T2w images for comparison purposes. Clinical interpretation of CDI^s, T2w, ADC, and K^{trans} was conducted by an expert radiologist with over 20 years of experience interpreting prostate MRI.

3.8 Experimental Results

3.8.1 Receiver Operating Characteristic Curve Analysis

Figure 3.4 shows the ROC curves for studying the performance of CDI^s SI, T2w SI, ADC values, and K^{trans} values for delineating csPCa tissue and insPCa tissue from healthy tissue. A number of observations can be made from this ROC analysis. First, it can be observed that CDI^s can achieve higher AUC for delineating between PCa tissue and healthy tissue (left plot in Figure 3.4) when compared to the standard MRI techniques, with ADC values achieving the next highest AUC (lower by 0.021 when compared to CDI^s_{MB}, $p < 0.0001$). However, no significant difference is observed between ADC and CDI^s_{MG} ($p = 0.1762$),

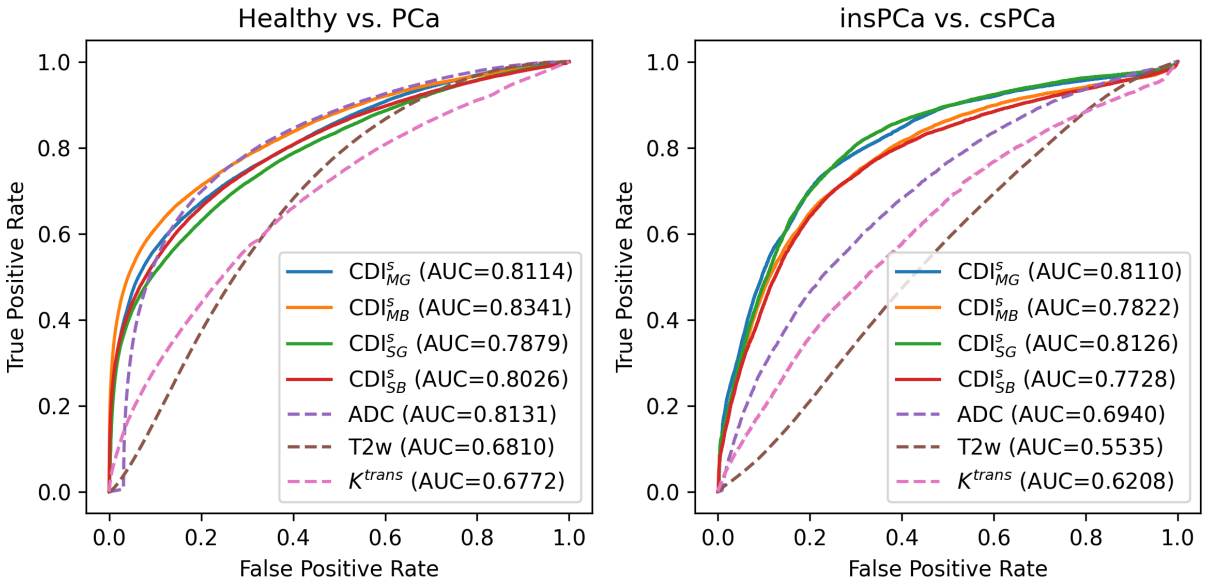

Figure 3.4: ROC curves for studying the performance of CDI^s SI, T2w SI, ADC values, and K^{trans} values for delineating csPCa tissue and insPCa tissue from healthy tissue. MG=(MEM, Gaussian filter), MB=(MEM, bilateral filter), SG=(SEM, Gaussian filter), SB=(SEM, bilateral filter).

and CDI^s_{SG} achieves significantly lower AUC than ADC and ($p < 0.0001$). T2w and K^{trans} values achieve significantly lower AUC compared to the other techniques (lower by as much as 0.1569 when compared to CDI^s_{MB}, $p < 0.0001$). CDI^s_{MB} achieves the highest AUC among the different forms of CDI^s (0.8341, 95% CI [0.8316, 0.8365]) and significantly outperforms CDI^s_{MG} (+0.0227 AUC, $p < 0.0001$), CDI^s_{SG} (+0.0462 AUC, $p < 0.0001$), and CDI^s_{SB} (+0.0315 AUC, $p < 0.0001$). Other relevant statistics are provided in Tables 3.2 and 3.3.

Considering clinical significance, it can be observed that CDI^s discriminates csPCa and insPCa tissue (right plot in Figure 3.4) more effectively than standard MRI techniques, with ADC, T2w, and K^{trans} AUC being as much as 0.1186 ($p < 0.0001$), 0.2591 ($p < 0.0001$), and 0.1918 ($p < 0.0001$) lower when compared to CDI^s_{SG}, respectively. CDI^s_{SG} achieves the highest AUC (0.8126, 95% CI [0.8078, 0.8173]) and significantly outperforms CDI^s_{MB} (+0.0304 AUC, $p < 0.0001$) and CDI^s_{SB} (+0.0398 AUC, $p < 0.0001$). No significant difference is observed between CDI^s_{SG} and CDI^s_{MG} ($p = 0.1346$). Other relevant statistics are provided in Tables 3.2 and 3.4.

Table 3.2: AUC values and 95% confidence intervals of different MRI modalities.MG=(MEM, Gaussian filter), MB=(MEM, bilateral filter), SG=(SEM, Gaussian filter), SB=(SEM, bilateral filter). The highest AUC in each column is highlighted in **bold**.

	AUC [95% CI]	
CDI^s Variant	**Healthy vs. PCa**	**insPCa vs. csPCa**
CDI^s_{MG}	0.8114 [0.8090, 0.8139]	0.8110 [0.8062, 0.8158]
CDI^s_{MB}	**0.8341** [0.8316, 0.8365]	0.7822 [0.7771, 0.7874]
CDI^s_{SG}	0.7879 [0.7853, 0.7906]	**0.8126** [0.8078, 0.8173]
CDI^s_{SB}	0.8026 [0.8000, 0.8052]	0.7728 [0.7675, 0.7780]
ADC	0.8131 [0.8109, 0.8152]	0.6940 [0.6884, 0.6996]
T2w	0.6810 [0.6804, 0.6817]	0.5535 [0.5519, 0.5551]
K^{trans}	0.6772 [0.6749, 0.6794]	0.6208 [0.6162, 0.6255]

Table 3.3: p-values of differences in healthy vs. PCa AUC between different MRI modalities. Values marked with * were obtained using unpaired z-tests, while all other values were obtained via paired z-tests. MG=(MEM, Gaussian filter), MB=(MEM, bilateral filter), SG=(SEM, Gaussian filter), SB=(SEM, bilateral filter).

	$\mathbf{CDI^s_{MB}}$	$\mathbf{CDI^s_{SG}}$	$\mathbf{CDI^s_{SB}}$	**ADC**	**T2w**	$\mathbf{K^{trans}}$
$\mathbf{CDI^s_{MG}}$	< 0.0001	< 0.0001	< 0.0001	0.1762	< 0.0001*	< 0.0001*
$\mathbf{CDI^s_{MB}}$	-	< 0.0001	< 0.0001	< 0.0001	< 0.0001*	< 0.0001*
$\mathbf{CDI^s_{SG}}$	-	-	< 0.0001	< 0.0001	< 0.0001*	< 0.0001*
$\mathbf{CDI^s_{SB}}$	-	-	-	< 0.0001	< 0.0001*	< 0.0001*
ADC	-	-	-	-	< 0.0001*	< 0.0001*
T2w	-	-	-	-	-	0.0010*

3.8.2 Clinical Interpretation

Figure 3.5(a,b) shows a T2w image and overlays of ADC, K^{trans}, and CDI^s for two patient cases with csPCa in the PZ. In Figure 3.5(a), it can be observed that T2w shows no contrast between csPCa tissue and healthy tissue, while K^{trans} exhibits strong contrast for a smaller portion within the csPCa tumour. ADC shows good contrast between the csPCa tumour and some of the surrounding healthy tissue, but exhibits ADC values similar to the tumour in different small regions within the TZ, including an adjacent region above the

Table 3.4: p-values of differences in insPCa vs. csPCa AUC between different MRI modalities. Values marked with * were obtained using unpaired z-tests, while all other values were obtained via paired z-tests. MG=(MEM, Gaussian filter), MB=(MEM, bilateral filter), SG=(SEM, Gaussian filter), SB=(SEM, bilateral filter).

	$\mathbf{CDI^s_{MB}}$	$\mathbf{CDI^s_{SG}}$	$\mathbf{CDI^s_{SB}}$	**ADC**	**T2w**	$\mathbf{K^{trans}}$
$\mathbf{CDI^s_{MG}}$	< 0.0001	0.1346	< 0.0001	< 0.0001	< 0.0001*	< 0.0001*
$\mathbf{CDI^s_{MB}}$	-	< 0.0001	< 0.0001	< 0.0001	< 0.0001*	< 0.0001*
$\mathbf{CDI^s_{SG}}$	-	-	< 0.0001	< 0.0001	< 0.0001*	< 0.0001*
$\mathbf{CDI^s_{SB}}$	-	-	-	< 0.0001	< 0.0001*	< 0.0001*
ADC	-	-	-	-	< 0.0001*	< 0.0001*
T2w	-	-	-	-	-	< 0.0001*

tumour that makes it indistinguishable from the tumour itself. CDI^s shows strong contrast for the entire csPCa tumour from the rest of the healthy tissue.

In Figure 3.5(b), it can be observed that T2w and K^{trans} show poor contrast between csPCa tissue and healthy tissue. ADC shows good contrast between the csPCa tumour and surrounding healthy tissue, but exhibits ADC values similar to the tumour in another small region within the PZ that was not identified as PCa tissue. CDI^s shows strong contrast for between the csPCa tumour from the rest of the healthy tissue.

Figure 3.5(c) shows a T2w image overlays of ADC, K^{trans}, and CDI^s for a patient case with csPCa in the TZ. It can be observed that T2w shows no contrast between csPCa tissue and healthy tissue, while K^{trans}, ADC, CDI^s all exhibit strong contrast between csPCa tissue and healthy tissue.

Figure 3.5(d) shows a T2w image and overlays of ADC, K^{trans}, and CDI^s for a patient with csPCa in the PZ and insPCa in the TZ. It can be observed that T2w shows no contrast between the csPCa tumour and healthy tissue, and poor contrast between the insPCa tumour and healthy tissue. K^{trans} exhibits no contrast between the csPCa tumour and healthy tissue, and strong contrast for a small portion of the insPCa tumour. ADC shows strong contrast between the csPCa tumour and surrounding healthy tissue and good contrast between the insPCa tumour and surrounding healthy tissue. However, ADC exhibits similar values for both the csPCa and insPCa tumours, as well as ADC values similar to both tumours in small regions within the TZ that were not identified as PCa tissue. CDI^s show the strongest contrast between the csPCa tumour and healthy tissue amongst the techniques, and shows good contrast between the insPCa tumour and healthy

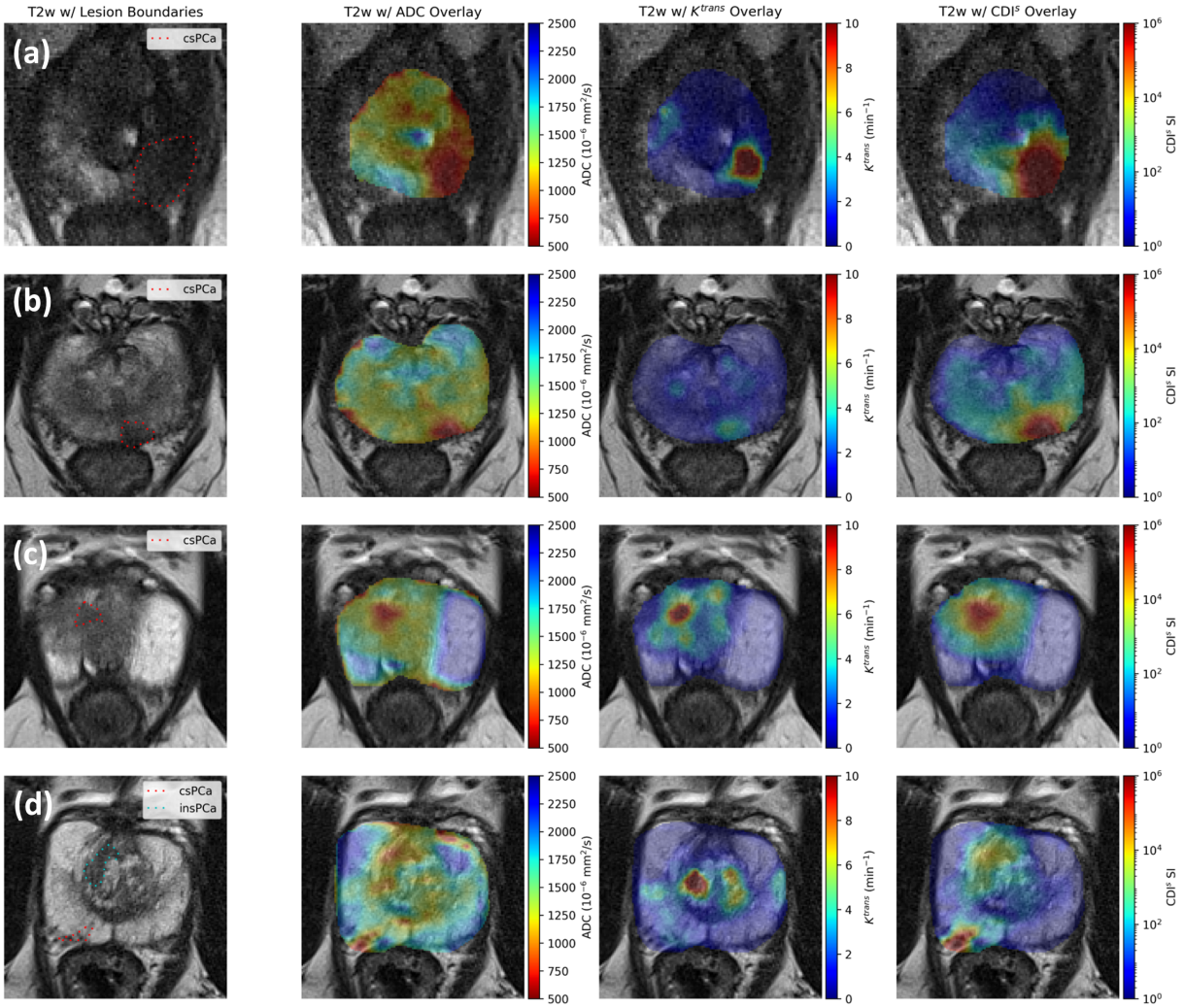

Figure 3.5: T2w images with overlays of le sion boundaries, ADC, K^{trans}, and CDI^s for four patient cases. **(a,b)** Two patients with csPCa in the PZ. **(c)** A patient with csPCa in the TZ. **(d)** A patient with csPCa in the PZ and insPCa in the TZ.

tissue. Furthermore,CDI^s provides greater contrast between the csPCa tumour and insPCa tumour than ADC.

3.9 Summary

A novel computed MRI modality called CDI^s was developed as an extension of CDI. CDI^s incorporates both native and synthetic DWI images through a weighted signal mixing procedure, and enforces a prior on the signal values using a linear or non-linear filter. CDI^s SI is calibrated by centering its log-normal distribution at $\sim$0 per-patient, thereby reducing inter- and intra-patient variability as well as variability due to differences in MRI hardware or protocols.

Quantitative results from a ROC curve analysis, presented in Figure 3.4, show that CDI^s can achieve a 0.021 increase in AUC over standard MRI modalities for discrimination of healthy tissue and PCa tissue, which may improve PCa screening. More significantly, a 0.1186 increase in AUC for discrimination of insPCa tissue and csPCa tissue was observed over standard MRI modalities, indicating considerable promise for CDI^s to be used as a tool for PCa risk assessment.

Clinical interpretation of CDI^s images by an experienced radiologist shows that CDI^s produces superior contrast between healthy tissue and PCa when compared to T2w and K^{trans}, and provides better contrast between insPCa and csPCa than ADC. This may aid radiologists in identifying PCa and distinuishing between insPCa and csPCa, potentially providing increased performance in PCa screening, biopsy guidance, and risk assessment.

Although these results are promising, the various parameters used to compute CDI^s in this chapter were hand-selected through a time-consuming "guess and check" procedure. Moreover, despite touting their potential benefits, the experiments in this chapter did not make use of the weighting parameters $\boldsymbol{\rho}$, instead opting for uniform weighting as in CDI. Overall, there is no clear procedure for choosing appropriate parameters, meaning that hand-selected parameters are unlikely to be optimal. This raises an interesting question: can appropriate CDI^s parameters be selected automatically? Chapter 4 will explore this question in more detail.

Chapter 4

Parameter Optimization

In this chapter, a gradient-based optimization framework is proposed through which the parameters of CDI^s may be automatically optimized for particular tasks; in particular, distinguishing PCa tissue from healthy tissue and delineation of insPCa and csPCa. To do so, several conceptual and numerical challenges must be overcome: how to automatically select b-values (Section 4.2), how to control numerical overflow (Section 4.3), how to optimize the prior PDF (Section 4.4), and how to define an appropriate optimization objective (Section 4.5). Once these challenges are addressed, Section 4.6 describes the experimental setup used to evaluate the optimization framework, with experimental results presented in Section 4.7. A brief summary is then provided in Section 4.8.

4.1 Problem Formulation

While a number of parameters are involved in the calculation of CDI^s, Chapter 3 did not discuss how to select these parameters. The particular choice of parameters can have a profound effect on the efficacy of CDI^s, yet in its original form these parameters must be specified by intuition or guess-and-check. For example, specifying low weights for clinically important b-values might drastically reduce performance. Conversely, one might choose higher weights for particular b-values based on evidence of optimal b-values from literature [68, 69], but there is no reason to conclude that such a weighting would be optimal. Manually choosing good parameters is also time-consuming, and so it is desirable to determine a method to optimize these parameters for the task at hand.

Examining the definition of CDI^s in Equation (3.5), it may be observed that CDI^s SI is

differentiable with respect to many of the parameters used to compute it. As such, gradient-based optimization techniques may theoretically be used to optimize said parameters for a particular task. However, in practice, several challenges remain which must be addressed in order to apply gradient-based techniques:

1. While parameters such as the weighting exponents and PDF parameters may be learned, the selection of b-values is not a differentiable operation. As such, b-value selection must be handled differently.

2. DWI SI has arbitrary units and typically has values on the order of 100s to 1000s. As a result, the mixing function M (Equation (3.1)) can have extremely large values before calibration is applied, and can grow unbounded as the number of b-values increases. This raises concerns regarding numerical stability, as intermediate calculations can quickly reach the limits of floating-point representation and cause the optimization to collapse.

3. The prior PDF f can take many forms which cannot all be accounted for in an optimization problem. As such, a suitable parameterized PDF must be defined which is sufficiently general and computationally tractable.

4. To effectively optimize CDI^s, an appropriate objective function must be specified which directly or indirectly reflects the desired performance metrics. Most notably, AUC was explored in Chapter 3, but AUC provides no useful gradient. Thus, AUC may not be used as the objective function directly, and an appropriate differentiable proxy must be identified.

Once the challenges above have been overcome, gradient-based optimization of CDI^s becomes relatively straightforward through the use of automatic differentiation [76], which allows for gradients to be computed automatically without being computed in closed-form by the user.

4.2 b-value Selection

While direct selection of b-values is not a differentiable operation, it is possible to indirectly select b-values using a convenient mathematical equivalence. Consider the general case where N b-values $\mathbf{b} = \begin{bmatrix} b_1 & \cdots & b_N \end{bmatrix}^\top$ are used to compute CDI^s, as defined in Equation (3.5). Now, suppose the k^{th} b-value is to be removed from the formulation. As long as

correspondences between $\mathbf{b}$ and $\boldsymbol{\rho}$ are preserved, the particular ordering of these vectors is arbitrary. Removal of the k^{th} b-value is therefore equivalent to removal of the 1^{st} b-value via simple reordering, and so without loss of generality the latter case is considered.

Removal of the 1^{st} b-value yields $\mathbf{b}' = \begin{bmatrix} b_2 & \cdots & b_N \end{bmatrix}^\top$ and $\boldsymbol{\rho}' = \begin{bmatrix} \rho_2 & \cdots & \rho_N \end{bmatrix}^\top$. Providing these modified vectors to the signal mixing function M, defined in Equation (3.1), yields

$$M(\mathbf{x}; \mathbf{b}', \boldsymbol{\rho}') = \prod_{i=2}^{N} \tilde{S}_{b_i}(\mathbf{x})^{\rho_i} = \tilde{S}_{b_2}(\mathbf{x})^{\rho_2} \tilde{S}_{b_3}(\mathbf{x})^{\rho_3} \cdots \tilde{S}_{b_N}(\mathbf{x})^{\rho_N} \tag{4.1}$$

Next, consider the case where $\mathbf{b}$ is left unchanged and $\rho_1 = 0$. The resulting exponent vector is $\boldsymbol{\rho}_0 = \begin{bmatrix} 0 & \rho_2 & \cdots & \rho_N \end{bmatrix}^\top$, and M then yields

$$M(\mathbf{x}; \mathbf{b}, \boldsymbol{\rho}_0) = \prod_{i=1}^{N} \tilde{S}_{b_i}(\mathbf{x})^{\rho_i} = \tilde{S}_{b_1}(\mathbf{x})^0 \tilde{S}_{b_2}(\mathbf{x})^{\rho_2} \tilde{S}_{b_3}(\mathbf{x})^{\rho_3} \cdots \tilde{S}_{b_N}(\mathbf{x})^{\rho_N} = M(\mathbf{x}; \mathbf{b}', \boldsymbol{\rho}') \tag{4.2}$$

since $\tilde{S}_{b_1}(\mathbf{x})^0 = 1$. Removal of the k^{th} b-value is therefore equivalent to setting $\rho_k = 0$.

Based on the equivalence above, one may conclude that selection of particular b-values is not strictly required. Instead, many uniformly-spaced b-values may be initially selected, and gradient-based optimization of $\boldsymbol{\rho}$ can indirectly act to remove b-values which are harmful to the performance of CDI$^{\text{s}}$ by setting their corresponding exponents to 0. By doing so, b-value selection is reduced to the selection of an interval $[b_a, b_b]$ from which N uniformly-spaced b-values are sampled. b_a, b_b, and N are then hyperparameters of the optimization problem.

4.3 Numerical Stability

DWI SI is measured in arbitrary units, and its scale may differ significantly between different MRI scanners. However, it is generally the case that $\tilde{S}_b >> 1$ for most reasonable b-values, with typical SIs being on the order of 100s to 1000s. Since the mixing function M combines these signals multiplicatively, the resulting values can become extremely large. This is particularly problematic given the b-value selection approach proposed in Section 4.2, where a large number of b-values are initially selected.

To address this issue, a method to constrain the possible output values of M is required. Examining Equation (3.1), it may be noted that M has a form similar to that of the

weighted geometric mean. Formally, the weighted geometric mean of positive values $\mathbf{v} = \begin{bmatrix} v_1 & \cdots & v_N \end{bmatrix}^\top$ with non-negative weights $\mathbf{w} = \begin{bmatrix} w_1 & \cdots & w_N \end{bmatrix}^\top$ is given by

$$\mathrm{GM}(\mathbf{v}; \mathbf{w}) = \prod_{i=1}^{N} v_i^{w_i / \sum_{j=1}^{N} w_j} \tag{4.3}$$

where $\sum_{j=1}^{N} w_j > 0$. It can be seen that this definition is quite similar to that of $M(\mathbf{x}; \mathbf{b}, \boldsymbol{\rho})$, but with an added division of the exponents by $\sum_{j=1}^{N} w_j$. Indeed, if a vector of signal values $\mathbf{d} = \begin{bmatrix} S_{b_1}(\mathbf{x}) & \cdots & S_{b_N}(\mathbf{x}) \end{bmatrix}^\top$ is defined, then

$$\mathrm{GM}(\mathbf{d}; \boldsymbol{\rho}) = \prod_{i=1}^{N} \tilde{S}_{b_i}(\mathbf{x})^{\rho_i / \sum_{j=1}^{N} \rho_j} = M(\mathbf{x}; \mathbf{b}, \boldsymbol{\rho})^{1 / \sum_{j=1}^{N} \rho_j} \tag{4.4}$$

subject to conditions $\rho_i \geq 0 \ \forall \ \rho_i \in \boldsymbol{\rho}$ and $\sum_{i=1}^{N} \rho_i > 0$. If these conditions are met, the following inequality holds:

$$\min_i \tilde{S}_{b_i}(\mathbf{x}) \leq M(\mathbf{x}; \mathbf{b}, \boldsymbol{\rho})^{1 / \sum_{j=1}^{N} \rho_j} \leq \max_i \tilde{S}_{b_i}(\mathbf{x}) \tag{4.5}$$

As such, regardless of the number of b-values used or the magnitudes of the exponents, the weighted geometric mean of the SIs is always bounded by the smallest and largest SIs, thereby preventing overflow in calculations.

While the result above is mathematically convenient, it overlooks the fact that $\rho_i \in \mathbb{R}$ in general and therefore may be negative. Indeed, during optimization, no constraint exists that would prevent negative exponents from appearing. To provide such a constraint, the exponents may be parameterized by their logarithm:

$$\tilde{\rho}_i = e^{\rho_i} \tag{4.6}$$

Under this parameterization, the scaled exponents may be conveniently expressed via the softmax function $\mathbf{s}$:

$$\frac{\tilde{\rho}_i}{\sum_{j=1}^{N} \tilde{\rho}_j} = \frac{e^{\rho_i}}{\sum_{j=1}^{N} e^{\rho_j}} = \mathbf{s}(\boldsymbol{\rho})_i \tag{4.7}$$

Applying the softmax function to $\boldsymbol{\rho}$ ensures that all exponents are positive and that they sum to 1, making it a valid way to produce a weighted geometric mean.

One final consideration is the purpose of the weighting exponents; they are meant to provide a mechanism by which signals at different b-values may be individually weighted

to optimize CDI^s SI for a particular task. However, when the exponents are parameterized via the softmax function, they must sum to 1 and are no longer independent, meaning that an increase in one exponent must correspond to a decrease in one or more of the other exponents. This limits the search space significantly, as in the general case the exponents can be any real number. For example, if $N = 100$, the exponents will have a mean of $1/N = 0.01$, and $|\mathbf{s}(\boldsymbol{\rho})_i - \mathbf{s}(\boldsymbol{\rho})_j| \leq 1 \; \forall \; i, j \in \{1, \ldots, N\}$. To provide additional flexibility to the optimization problem, a gain $K \geq 1$ is introduced, yielding

$$M(\mathbf{x}; \mathbf{b}, K\mathbf{s}(\boldsymbol{\rho})) = \prod_{i=1}^{N} \tilde{S}_{b_i}(\mathbf{x})^{K\mathbf{s}(\boldsymbol{\rho})_i} \tag{4.8}$$

By adding this gain term, mixed signals are still bounded as in Equation (4.5), but now

$$\min_i \tilde{S}_{b_i}(\mathbf{x})^K \leq M(\mathbf{x}; \mathbf{b}, K\mathbf{s}(\boldsymbol{\rho})) \leq \max_i \tilde{S}_{b_i}(\mathbf{x})^K \tag{4.9}$$

or in log space,

$$K \min_i \ln \tilde{S}_{b_i}(\mathbf{x}) \leq \ln M(\mathbf{x}; \mathbf{b}, K\mathbf{s}(\boldsymbol{\rho})) = K \ln M(\mathbf{x}; \mathbf{b}, \mathbf{s}(\boldsymbol{\rho})) \leq K \max_i \ln \tilde{S}_{b_i}(\mathbf{x}) \tag{4.10}$$

Hence, the gain term scales the log-M distribution by K, providing additional flexibility in the optimization problem. Equivalently, the gain may be thought of as modifying the upper bound on the exponents:

$$0 \leq K\mathbf{s}(\boldsymbol{\rho})_i \leq K \quad \forall \; i \in \{1, \ldots N\} \tag{4.11}$$

4.4 Prior Optimization

Optimization of the prior PDF (f) must also be considered in the proposed optimization framework, as hand-design of a prior may not be optimal. In Section 3.3, two potential filters were discussed: the Gaussian filter and the bilateral filter. However, in Section 3.8, it was found that the linear Gaussian filter often outperformed the nonlinear bilateral filter. Additionally, linear filters are significantly more efficient, which is beneficial in the case of iterative optimization. As such, a trainable linear filter is adopted as the prior for the CDI^s optimization problem.

Intuitively, the simplest way to define a trainable linear filter is to set its spatial size (in voxels) to that of $V(\mathbf{x})$ and optimize the values of its kernel tensor directly. Given that DWI consists of discrete voxels, this kernel tensor has a fixed size for a given voxel

resolution. However, it is still required that f be a valid (discrete) distribution over $V(\mathbf{x})$, and as such the kernel values must be non-negative and sum to 1. To achieve this, the same softmax approach explored for constraining $\boldsymbol{\rho}$ (Section 4.3) may be used; each kernel value $\tilde{k}_{xyz}$ is parameterized via its the logarithm and normalized by the sum of all kernel values. Formally,

$$\tilde{k}_{xyz} = \frac{e^{k_{xyz}}}{\sum_{x',y',z'} e^{k_{x'y'z'}}} = \mathbf{s}(\mathbf{k})_{xyz} \tag{4.12}$$

where x, y, z represent spatial indices of the kernel and $\mathbf{k}$ is a flattened vector representation of the kernel parameters.

With the changes above, the form of CDI^s which will be used for optimization is

$$C_s(\mathbf{x}; \mathbf{b}, K\mathbf{s}(\boldsymbol{\rho}), \mathbf{s}(\mathbf{k})) = \frac{1}{Z} \sum_{\mathbf{x}' \in V(\mathbf{x})} M(\mathbf{x}'; \mathbf{b}, K\mathbf{s}(\boldsymbol{\rho})) f\left(\mathbf{x}'; \mathbf{s}(\mathbf{k})\right) \tag{4.13}$$

where $f(\mathbf{x}; \mathbf{s}(\mathbf{k}))$ is a linear filter with kernel $\mathbf{s}(\mathbf{k})$, and all other symbols are as previously defined.

4.5 Loss Functions

In Chapter 3, AUC was used as the evaluation metric for CDI^s and standard MRI modalities. However, AUC is a discontinuous function with a gradient of 0 almost everywhere, and hence it cannot be used as an objective function for gradient-based optimization.

Mathematically, AUC may be defined via element-wise comparisons of positive samples $p \in \mathcal{D}_p$ and negative samples $n \in \mathcal{D}_n$. For example, positive samples may be CDI^s values within csPCa tumours while negative samples may be CDI^s values within insPCa tumours. AUC may be thought of as the probability that a randomly-selected positive sample is larger than a randomly-selected negative sample, and can be expressed via the heaviside step function H:

$$H(x) = \begin{cases} 1, & x > 0 \\ \frac{1}{2}, & x = 0 \\ 0, & x < 0 \end{cases} \tag{4.14}$$

$$\text{AUC}(\mathcal{D}_n, \mathcal{D}_p) = \frac{\sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} H(p - n)}{|\mathcal{D}_n||\mathcal{D}_p|} \tag{4.15}$$

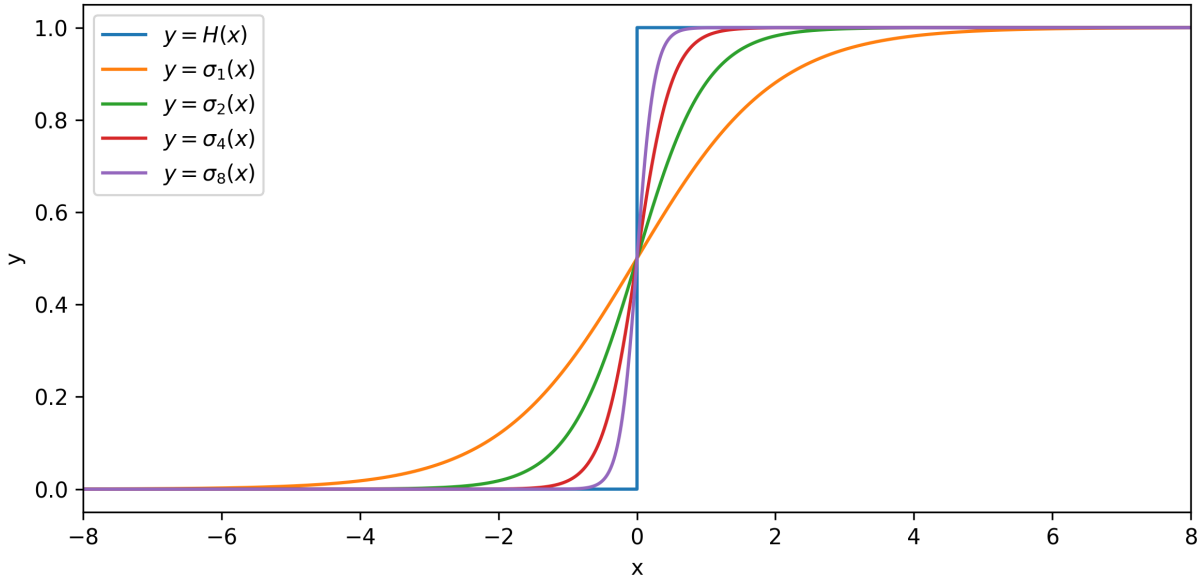

Figure 4.1: Illustration of the heaviside function and various sigmoid functions. As β increases, $\sigma_\beta(x)$ becomes closer to $H(x)$.

The use of H is what causes AUC to be a poor objective function, as H provides no useful gradient. Given this, an intuitive approach to defining a differentiable proxy for AUC is to replace H with a similar smooth function. One such function is the sigmoid function, parameterized by $\beta > 0$:

$$\sigma_\beta(x) = \frac{1}{1 + e^{-\beta x}} \tag{4.16}$$

Figure 4.1 shows plots of H and σ_β with various values of β. As β increases, σ_β becomes a better approximation of H, and in the limiting case

$$\lim_{\beta \to \infty} \sigma_\beta(x) = H(x) \tag{4.17}$$

By replacing H with σ_β in Equation (4.15), the so-called soft-AUC is obtained [77]:

$$\mathrm{AUC}_{\mathrm{soft}}(\mathcal{D}_n, \mathcal{D}_p; \beta) = \frac{\sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} \sigma_\beta(p - n)}{|\mathcal{D}_n||\mathcal{D}_p|} \tag{4.18}$$

This is a smooth differentiable function bounded on $[0, 1]$. AUC may then be optimized

by minimizing the soft-AUC loss $\mathcal{L}_{\text{soft}}$, defined as the complement of soft-AUC (see Appendix A.2):

$$\mathcal{L}_{\text{soft}}(\mathcal{D}_n, \mathcal{D}_p; \beta) = 1 - \text{AUC}_{\text{soft}}(\mathcal{D}_n, \mathcal{D}_p; \beta) = \frac{\sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} \sigma_\beta(n - p)}{|\mathcal{D}_n||\mathcal{D}_p|} \tag{4.19}$$

Although increasing β improves the approximation of AUC, it also has drawbacks. First, as illustrated in Figure 4.1, large values of β cause $\sigma_\beta(x)$ to saturate closer to $x = 0$, producing small gradients which may slow optimization. Moreover, the derivative of $\mathcal{L}_{\text{soft}}$ with respect to $d = n - p$ is scaled by β and may become steep when β is large:

$$\frac{\partial \mathcal{L}_{\text{soft}}}{\partial d} = \frac{\partial}{\partial d} \frac{\sigma_\beta(d)}{|\mathcal{D}_n||\mathcal{D}_p|} = \beta \frac{\sigma_\beta(d)(1 - \sigma_\beta(d))}{|\mathcal{D}_n||\mathcal{D}_p|} \tag{4.20}$$

This problem was explored by Yan et al. [78], where it was found that $\beta > 2$ can lead to numerical problems which prevent successful optimization. Instead, they note that AUC is exactly the normalized Wilcoxon-Mann-Whitney (WMW) statistic and propose an alternative loss function as a proxy for the WMW statistic:

$$\mathcal{L}_{\text{WMW}}(\mathcal{D}_n, \mathcal{D}_p; \varepsilon, k) = \sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} \begin{cases} (-(p - n - \varepsilon))^k, & p - n < \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{4.21}$$

where $k > 1$. $\varepsilon \in [0, 1]$ acts as a margin to enforce a minimum separation between n and p. However, this margin is difficult to specify – the original interval $[0, 1]$ considers the case of $n, p \in [0, 1]$, which is common for classifiers but is not the case for CDI^s.

Calders and Jaroszewicz [77] proposed a different approximation of H using a polynomial (Chebyshev) approximation of degree d:

$$H(x) \approx \sum_{k=0}^{d} c_k x^k \tag{4.22}$$

However, as shown in Figure 4.2, this approximation is limited to a particular interval, beyond which the approximation becomes inaccurate. Although any desired interval can be specified, for a fixed degree d the approximation becomes less accurate as the interval widens, particularly near $x = 0$. In the case of CDI^s, its high dynamic range necessitates a wide interval, meaning that a very high degree approximation would be required for a useful approximation of $H(x)$.

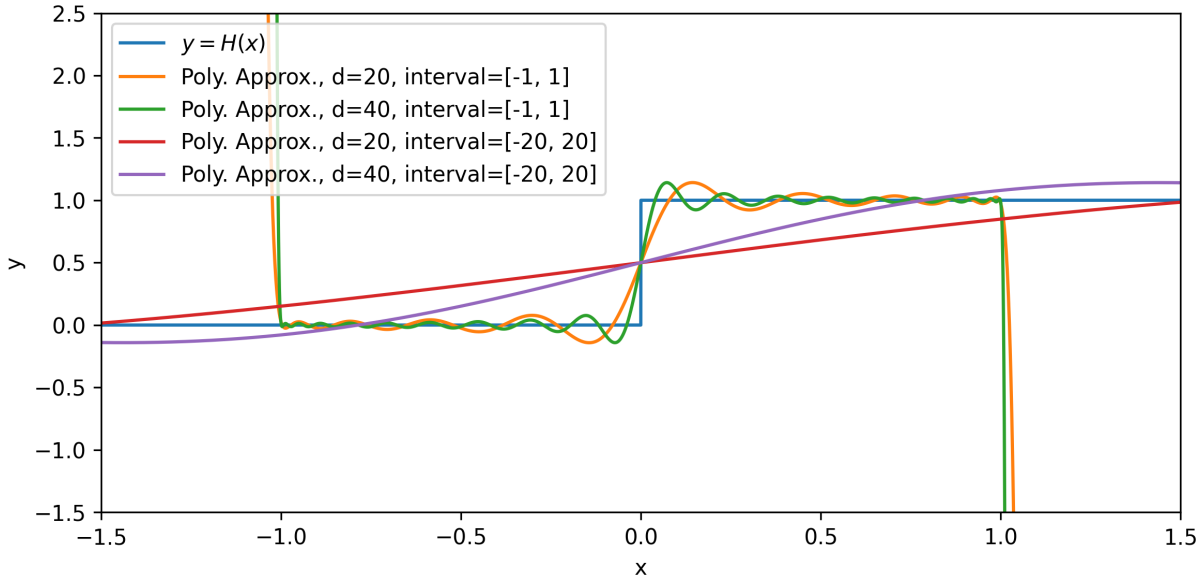

Figure 4.2: Illustration of the heaviside function and various polynomial approximations. As d increases, the approximation becomes closer to $H(x)$. As the interval widens, the approximation becomes less accurate, particularly near $x = 0$.

In initial experiments with CDI^s, it was found that both the soft-AUC and approximate WMW losses could be optimized, with the soft-AUC generally providing better performance. In particular, it was found that $\beta > 2$ did not pose any noticeable gradient issues (in contrast to the findings of Yan et al. [78]), and as such $\mathcal{L}_{\text{soft}}$ was adopted as the loss function of choice.

4.6 Experimental Setup

The numerically-stable formulation of CDI^s proposed in Equation (4.13) was used in all experiments, with 101 b-values evenly-spaced on the interval [0 s/mm^2, 5000 s/mm^2] (i.e., in steps of 50 s/mm^2). Both the MEM and SEM were explored for generating synthetic DWI acquisitions, and native acquisitions at b-values of 50 s/mm^2, 400 s/mm^2, and 800 s/mm^2 were also included in all experiments. A gain of $K = 10$ was used, and $V(\mathbf{x})$ was defined as a $18 \times 18 \times 6$ mm^3 volume centered at $\mathbf{x}$, with a corresponding trainable filter defined accordingly as a $9 \times 9 \times 3$ tensor. The cylindrical calibration mask described in Section 3.4

Table 4.1: Abbreviations of CDI^s variants, as used in Tables 4.2 to 4.4.

CDI^s Variant	**Decay Model**	**Filter**	**Loss Function**
MO1	MEM	Linear	$\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta)$
MO2	MEM	Linear	$\mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)$
MO3	MEM	Linear	$\frac{1}{2}\left[\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta) + \mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)\right]$
SO1	SEM	Linear	$\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta)$
SO2	SEM	Linear	$\mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)$
SO3	SEM	Linear	$\frac{1}{2}\left[\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta) + \mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)\right]$
MG	MEM	Gaussian	N/A
MB	MEM	Bilateral	N/A
SG	SEM	Gaussian	N/A
SB	SEM	Bilateral	N/A

was used in all experiments with a radius of $0.25 \min(H, W)$, where H and W are the height and width of the DWI slices, respectively.

For optimization, CDI^s values were divided into sets $\mathcal{H}$, $\mathcal{I}$, and $\mathcal{S}$ depending on whether they represent healthy tissue, insPCa tissue, or csPCa tissue, respectively. The soft-AUC loss (Equation (4.18)) with $\beta = 4$ was used as the loss function in all experiments, and CDI^s was optimized for three different tasks:

1. Discrimination of of healthy tissue and PCa tissue: $\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta)$.
2. Discrimination of insPCa tissue and csPCa tissue: $\mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)$.
3. Both of the tasks above: $\frac{1}{2}\left[\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta) + \mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)\right]$.

To distinguish between different experiments, Table 4.1 lists abbreviations and specifications for each variant of CDI^s examined in this chapter.

Optimization was performed via stochastic gradient descent (SGD) with momentum [79]. An initial learning rate of 0.1 and momentum of 0.9 were used in all experiments, and the learning rate was gradually decreased to 0.0001 over the course of training via cosine decay [80]. Optimization proceeded for 2500 epochs (32500 iterations) with a batch size of 16.

Due to the limited number of patients, dedicated training and evaluation splits are likely to be biased by the particular choice of splits. While a cross-validation technique

such as k-fold cross validation could be applied to address this, results across different folds cannot be combined due to differences in CDI^s SI distributions. Simply averaging AUC across folds may also be inaccurate because the number of samples differs (i.e., the number of patients is consistent across folds, but the number of healthy/insPCa/csPCa voxels differs). As such, in this thesis CDI^s is optimized and evaluated using all of the available data. While this may yield results which are biased high, it still illustrates the potential benefits of CDI^s optimization and allows for paired statistical tests to be performed.

To assess the statistical significance of differences in AUC values between different CDI^s variants, the estimates of AUC variance and covariance proposed by DeLong et al. [74] are adopted, with paired two-tailed z-tests performed in all cases. Comparisons with ADC, T2w, and K^{trans} were not performed, as such tests were performed in Chapter 3 and the purpose of these experiments is to examine differences between optimized CDI^s variants and the baseline hand-tuned CDI^s variants presented in Chapter 3. An unadjusted significance level of 0.05 was initially selected; however, since 39 tests are being performed (i.e., tests between every possible pairing of optimized CDI^s variants, as well as between pairs of optimized and baseline variants), the Bonferroni correction [75] was applied, yielding an adjusted significance level of $0.05/39 = 0.00\overline{128205}$.

To explore the effects of the various hyperparameters on the optimization problem, a number of experiments were performed in which a single hyperparameter was varied. In particular, the b-value interval $[b_a, b_b]$, soft-AUC loss parameter β, gain K, and size of the filter were all varied individually to examine their effect on CDI^s performance following optimization. All of these experiments were performed with the MEM and the insPCa vs. csPCa objective defined in Section 4.6 (i.e., $\mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)$). The parameters which were not varied in a particular experiment were fixed at $[b_a, b_b] = [0\ \text{s/mm}^2, 5000\ \text{s/mm}^2]$, $\beta = 4$, $K = 10$, and kernel size $7 \times 7 \times 3$.

The CDI^s models, loss functions, and training/evaluation framework were implemented using the PyTorch library [76].

4.7 Experimental Results

AUC values for discrimination of healthy tissue and PCa tissue and discrimination of insPCa and csPCa tissue are tabulated in Table 4.2, along with their corresponding 95% CIs. p-values of AUC differences are also provided in Tables 4.3 and 4.4.

Table 4.2: AUC values and 95% confidence intervals of different CDI^s variants. Averages and absolute differences in AUC values between the two tasks are also shown. The highest AUC in each column is highlighted in **bold**, as well as the lowest absolute AUC difference. CDI^s variants are as defined in Table 4.1.

	AUC [95% CI]			
CDI^s Variant	$\mathcal{H}$ vs. $\mathcal{I} \cup \mathcal{S}$	$\mathcal{I}$ **vs.** $\mathcal{S}$	**Avg.**	**Diff.**
MO1	**0.8446** [0.8423, 0.8469]	0.7713 [0.7662, 0.7764]	0.8079	0.0733
MO2	0.7860 [0.7834, 0.7886]	**0.8530** [0.8488, 0.8571]	0.8195	0.0670
MO3	0.8240 [0.8216, 0.8264]	0.8327 [0.8282, 0.8372]	**0.8284**	0.0087
SO1	0.8109 [0.8083, 0.8134]	0.7776 [0.7726, 0.7826]	0.7943	0.0333
SO2	0.7769 [0.7742, 0.7795]	0.8524 [0.8482, 0.8567]	0.8146	0.0755
SO3	0.8062 [0.8037, 0.8088]	0.8321 [0.8277, 0.8366]	0.8192	0.0259
MG	0.8114 [0.8090, 0.8139]	0.8110 [0.8062, 0.8158]	0.8112	**0.0004**
MB	0.8341 [0.8316, 0.8365]	0.7822 [0.7771, 0.7874]	0.8082	0.0519
SG	0.7879 [0.7853, 0.7906]	0.8126 [0.8078, 0.8173]	0.8003	0.0247
SB	0.8026 [0.8000, 0.8052]	0.7728 [0.7675, 0.7780]	0.7994	0.0065

4.7.1 Comparison with Baseline CDI^s Variants

Examining Tables 4.2 to 4.4, it is observed that optimization of CDI^s for a particular task yields significant improvement in AUC over the baseline CDI^s variants. For example, MO1 outperforms both MG (+0.0332 AUC, $p < 0.0001$) and MB (+0.0105 AUC, $p < 0.0001$) for discrimination of healthy and PCa tissue, while MO2 outperforms both MG (+0.0420 AUC, $p < 0.0001$) and MB (+0.0708 AUC, $p < 0.0001$) for discrimination of insPCa and csPCa tissue. Similarly, SO1 outperforms both SG (+0.0230 AUC, $p < 0.0001$) and SB (+0.0083 AUC, $p < 0.0001$) for discrimination of healthy and PCa tissue, while SO2 outperforms both SG (+0.0398 AUC, $p < 0.0001$) and SB (+0.0796 AUC, $p < 0.0001$) for discrimination of insPCa and csPCa tissue.

Overall, AUC gains of 0.0105 and 0.0404 are seen between the best optimized variants and best baseline variants for healthy vs. PCa and insPCa vs. csPCa discrimination, respectively. These results illustrate the efficacy of the proposed CDI^s optimization framework, particularly for discrimination of insPCa and csPCa tissue where the AUC gains obtained through optimization are considerable.

Table 4.3: p-values of differences in healthy vs. PCa AUC between different CDI^s variants. All values were obtained via paired z-tests, and the CDI^s variants are as defined in Table 4.1.

	CDI^s Variant					
	MO1	**MO2**	**MO3**	**SO1**	**SO2**	**SO3**
MO2	< 0.0001	-	-	-	-	-
MO3	< 0.0001	< 0.0001	-	-	-	-
SO1	< 0.0001	< 0.0001	< 0.0001	-	-	-
SO2	< 0.0001	< 0.0001	< 0.0001	< 0.0001	-	-
SO3	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	-
MG	< 0.0001	< 0.0001	< 0.0001	0.4873	< 0.0001	< 0.0001
MB	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SG	< 0.0001	0.0119	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SB	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Table 4.4: p-values of differences in insPCa vs. csPCa AUC between different CDI^s variants. All values were obtained via paired z-tests, and the CDI^s variants are as defined in Table 4.1.

	CDI^s Variant					
	MO1	**MO2**	**MO3**	**SO1**	**SO2**	**SO3**
MO2	< 0.0001	-	-	-	-	-
MO3	< 0.0001	< 0.0001	-	-	-	-
SO1	< 0.0001	< 0.0001	< 0.0001	-	-	-
SO2	< 0.0001	0.3747	< 0.0001	< 0.0001	-	-
SO3	< 0.0001	< 0.0001	0.4701	< 0.0001	< 0.0001	-
MG	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
MB	< 0.0001	< 0.0001	< 0.0001	0.0036	< 0.0001	< 0.0001
SG	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SB	0.3706	< 0.0001	< 0.0001	0.0006	< 0.0001	< 0.0001

4.7.2 Comparison of Decay Models

Considering the task of separating healthy tissue from PCa tissue (i.e., $\mathcal{H}$ vs. $\mathcal{I} \cup \mathcal{S}$), it can be observed that the MEM generally outperforms the SEM. Indeed, when the SEM

and MEM results for the same loss function are compared, the MEM achieves significantly higher AUC in all three cases ($p < 0.0001$), as shown in Tables 4.2 and 4.3. The highest AUC for this task (0.8446) is achieved using the MEM, which outperforms the best SEM AUC by 0.0337 ($p < 0.0001$).

When delineation of insPCa and csPCa is considered, the differences between the MEM and SEM are less obvious. Examining Tables 4.2 and 4.4, no significant differences in AUC are observed between MO2 and SO2 ($p = 0.3747$) or between MO3 and SO3 ($p = 0.4701$). Although SO1 significantly outperformed MO1 for this task (+0.0063 AUC, $p < 0.0001$), both SO1 and MO1 were optimized for a different task.

Based on the observations above, the MEM may be a better choice of decay model for the tasks and dataset explored in this thesis, as the MEM achieves significantly higher $\mathcal{H}$ vs. $\mathcal{I} \cup \mathcal{S}$ AUC than the SEM and achieves similar AUC for delineation of insPCa and csPCa. This echoes the findings of previous studies [48, 49, 53, 58], where the SEM was not found to be superior to the MEM. One possible reason for this is that the SEM's parameters are estimated from only 3 b-values which are all ≤ 800 s/mm^2, and as such the true decay curve may not be adequately sampled. Typically, studies which explore the SEM use more b-values (e.g., 4 [48, 53], sometimes 12 [49] or more) and include at least one b-value above 1000 s/mm^2 (e.g., 1500 s/mm^2 [53] or 2000 s/mm^2 [48, 49]). For the dataset explored in this thesis, the estimates of α are particularly noisy, which may hinder accurate extrapolation of the DWI signal.

4.7.3 Effect of Loss Function

As one might expect, the highest AUC values for a particular task are achieved when the chosen loss function matches the task. For example, the highest $\mathcal{H}$ vs. $\mathcal{I} \cup \mathcal{S}$ AUC (0.8446, 95% CI [0.8423, 0.8469]) is achieved when the loss function is $\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta)$, and the highest $\mathcal{I}$ vs. $\mathcal{S}$ AUC (0.8530, 95% CI [0.8488, 0.8571]) is achieved when the loss function is $\mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)$. However, a notable trade-off is observed; optimization for one task leads to a decrease in AUC for the other. This is reflected in the average AUC column of Table 4.2, where optimizing CDI^s for a single task yields a lower average than when CDI^s is jointly optimized for both tasks. Moreover, examining the AUC difference column in Table 4.2, it can be observed that joint optimization yields a better balance of the task-wise AUC values.

From a practical standpoint, the trade-off noted above may be of no consequence, as one may simply optimize CDI^s for each task separately and then use both variants at different times. For example, optimization with $\mathcal{L}_{\text{soft}}(\mathcal{H}, \mathcal{I} \cup \mathcal{S}; \beta)$ is most useful for PCa

screening or diagnosis. Once PCa has been confirmed, a variant of CDI^s optimized with $\mathcal{L}_{\text{soft}}(\mathcal{I}, \mathcal{S}; \beta)$ can then be used to aid in risk assessment and treatment planning.

4.7.4 Effect of b-value Interval

To explore the impact of the b-value interval, the upper bound of the b-value interval (b_b) was varied from 500 s/mm^2 to 16000 s/mm^2 while retaining a fixed lower bound of $b_a = 0$ s/mm^2 and fixed spacing of 50 s/mm^2 between consecutive b-values. However, the three native DWI acquisitions at $b = 50$ s/mm^2, $b = 400$ s/mm^2, and $b = 800$ s/mm^2 were always included.

As shown in Figure 4.3, the AUC increases with b_b but plateaus beyond $b_b = 2000$ s/mm^2, which is consistent with past studies of DWI for PCa imaging that found that b-values in the range of approximately 1000-2500 s/mm^2 are optimal for PCa detection and risk assessment [68, 69]. This result also aligns with the hypothesis that the inclusion of "extra" b-values which are not needed can be corrected by optimization of the exponents. This is further illustrated by examining plots of $\ln(K\mathbf{rho}(\boldsymbol{\rho})_i)$ vs. b_i, as shown in Figure 4.4. At low values of b_b, the optimization of $\boldsymbol{\rho}$ leads to a single b-value receiving a high weight, but reduced performance. As b_b increases, the exponents converge towards a smooth curve with a consistent shape, and high b-values are effectively ignored by making their exponents very small.

4.7.5 Effect of β

Figure 4.5 illustrates the impact β has on AUC. Small values of β (i.e., $\beta < 2$) yield poor approximations of the heaviside function, increasing the discordance between the optimization objective (maximization of soft-AUC) and the evaluation metric (AUC). As β increases, the objective and evaluation functions become more similar, and the soft-AUC loss therefore provides an optimization objective which better coincides with direct AUC optimization. Beyond $\beta = 2$, the two functions may be similar enough that further increase of β yields marginal or no improvement in performance. Interestingly, very high values of β do not seem to cause issues due saturation of σ_β.

4.7.6 Effect of Gain

Figure 4.6 illustrates how the exponent gain K influences AUC following optimization. As shown, AUC increases with gain and plateaus beyond $K = 10$. Since K directly

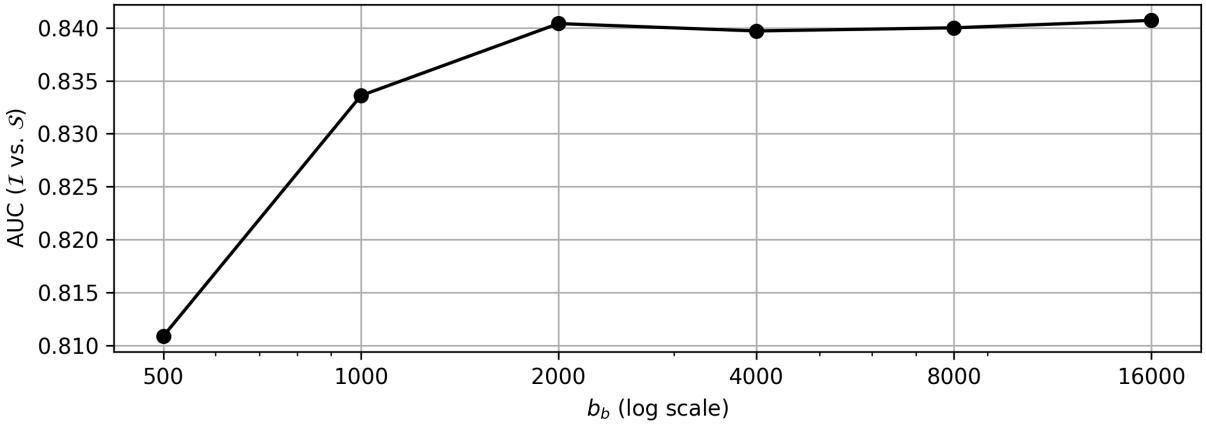

Figure 4.3: Effect of upper b-value bound b_b on AUC for delineation of insPCa and csPCa. When b_b is small, performance is markedly reduced. Performance increases with b_b and plateaus beyond $b_b = 2000$ s/mm^2.

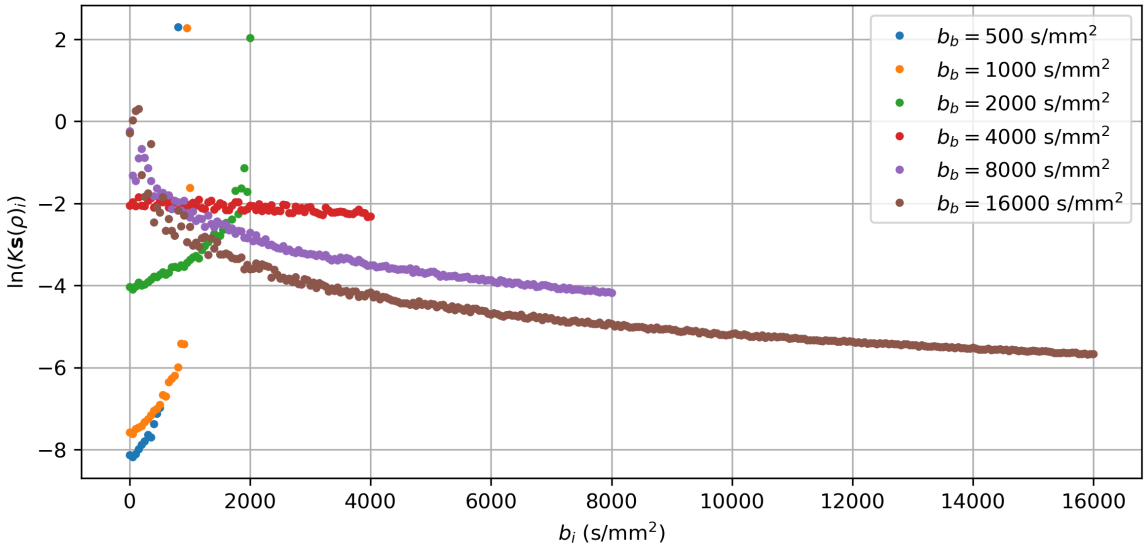

Figure 4.4: Plots of $\ln(K\mathbf{s}(\boldsymbol{\rho})_i)$ vs. b_i for various values of b_b. As b_b increases, the exponents converge towards a particular shape.

influences the range of possible exponents, greater gain may be thought of as providing greater flexibility in the optimization problem. However, once the gain is adequately large,

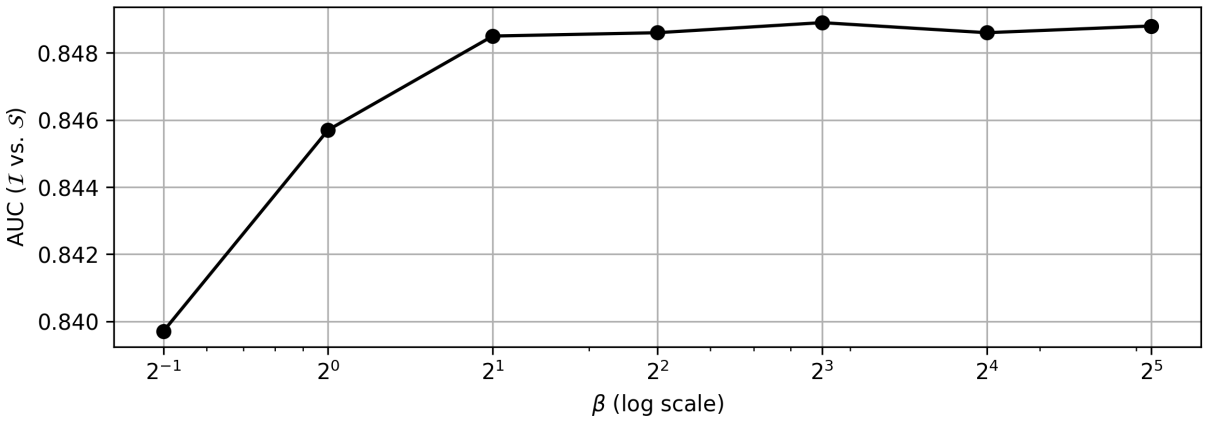

Figure 4.5: Effect of β on AUC for delineation of insPCa and csPCa. When β is small, performance is reduced. Performance increases with β and plateaus beyond $\beta = 2$.

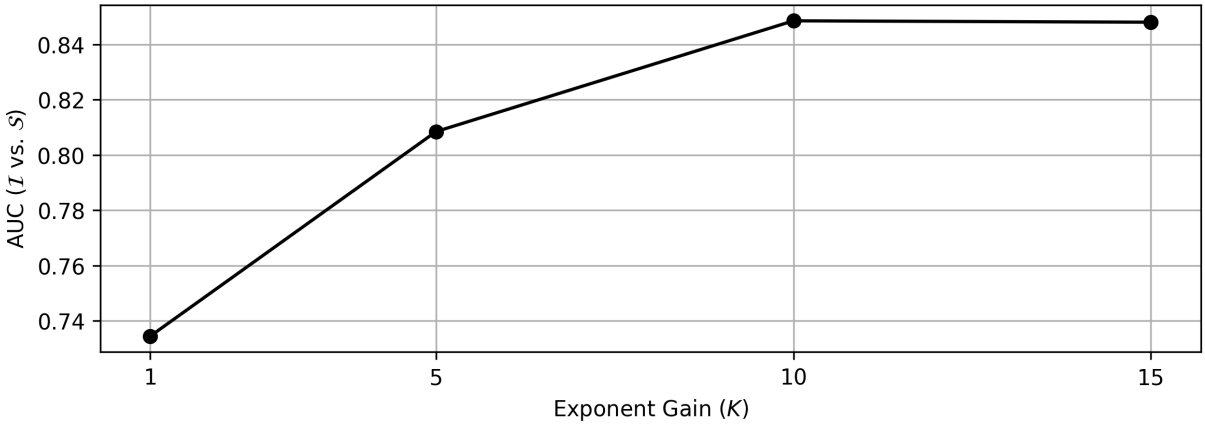

Figure 4.6: Effect of exponent gain K on AUC for delineation of insPCa and csPCa. Increasing the gain increases the range of possible exponents, yielding additional flexibility in the optimization problem and increased AUC by extension.

there is "enough" flexibility, and only marginal changes in AUC are observed beyond this point.

Notably, the gain also directly impacts the range of intermediate values produced by M, as formally described in Equation (4.9). As a result, if the gain is made too large, numerical overflow can still occur. Indeed, a gain of $K = 20$ was also tested and found

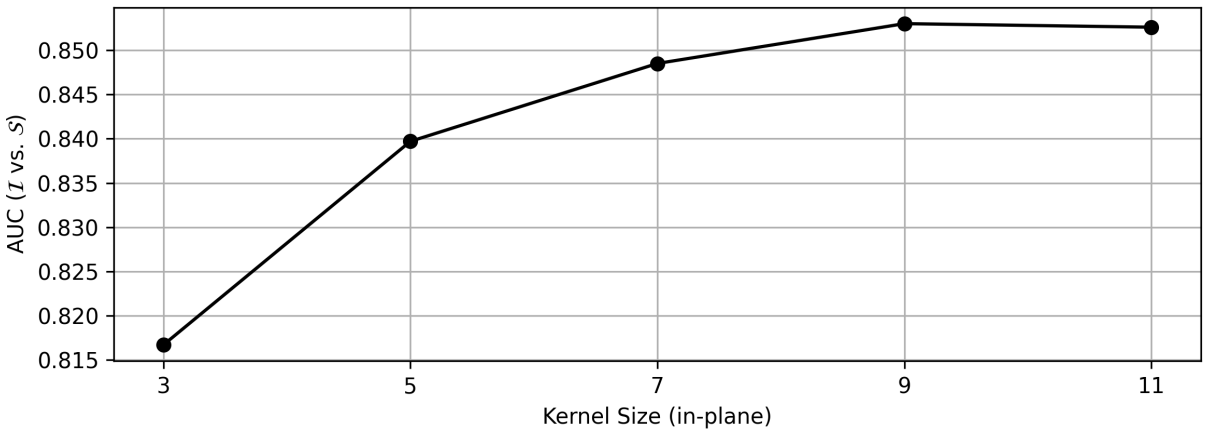

Figure 4.7: Effect of in-plane kernel size on AUC for delineation of insPCa and csPCa. When the size is small, performance is markedly reduced. Performance increases with kernel size and plateaus beyond 9.

to rapidly result in overflow. As such, the gain must be carefully chosen so as to balance flexibility with numerical limitations.

4.7.7 Effect of Kernel Size

For these experiments, the kernel size was set to $I \times I \times 3$, where I denotes the in-plane size and 3 is the number of slices (i.e., only the two slices immediately adjacent to the current slice are included by the filter). Figure 4.7 illustrates how AUC changes as the in-plane kernel size is varied from 3×3 to 11×11. As shown, smaller kernels yield poorer performance, and AUC plateaus as the kernel size increases beyond 9.

Intuitively, as the kernel size increases, voxels farther from the central voxel location $\mathbf{x}$ can influence the CDI SI at $\mathbf{x}$. However, this may not be desired, as beyond a certain distance the voxels in the neighbourhood of $\mathbf{x}$ may contribute spurious signals which should not logically influence the SI at $\mathbf{x}$. Optimization of the kernel's values can account for this; when the kernel is larger than necessary, it is still possible to have a kernel with a smaller *effective size* via optimization of the kernel values. Conversely, when the kernel is too small, it is not possible to make its effective size larger. Therefore, one can initially make the kernel larger than necessary, then reduce its size to improve computational efficiency.

4.8 Summary

In this chapter, the definition of CDI^s that was developed in Chapter 3 was adapted in order to allow gradient-based optimization to be applied to its parameters. In particular, the mixing of native and synthetic DWI signals was reformulated as a weighted geometric mean via constrained exponents, and a trainable linear filter was defined to act as the spatial prior. Furthermore, loss functions which approximate AUC were explored in order to provide an appropriate optimization objective.

Optimization experiments were performed using both the MEM and SEM for synthetic signal generation, as well as with three different loss functions based on the soft-AUC. Comparisons between optimized CDI^s and the baseline CDI^s variants presented in Chapter 3 show that optimization can improve AUC for discrimination of healthy tissue and PCa tissue by 0.0105, and can improve AUC for discrimination of insPCa tissue and csPCa tissue by 0.0404. Furthermore, comparisons between the MEM and SEM showed that the additional complexity of the SEM does not lead to performance gains, with the MEM equalling or exceeding the performance of the SEM in most cases. Experiments using different loss functions showed that a performance trade-off exists between healthy and PCa tissue discrimination and insPCa and csPCa tissue discrimination, and that the best results are obtained when the loss function matches the evaluation metric. Lastly, the impacts of the various hyperparameters on discrimination performance were explored, with the results indicating that hyperparameters may be made larger than necessary without harming performance.

Chapter 5

Conclusion

5.1 Summary of Thesis and Contributions

In this thesis a novel computed MRI modality called CDI^s was developed, as well as a framework for optimization of its parameters.

In Chapter 3, CDI^s was developed as an extension of CDI which incorporates synthetic DWI images. CDI^s was designed to be more flexible and consistent than CDI by incorporating synthetic DWI, weighted mixing of native and synthetic DWI, and per-patient calibration. Quantitative results from a ROC curve analysis, presented in Figure 3.4, show that CDI^s can achieve a 0.021 increase in AUC over standard MRI modalities for discrimination of healthy tissue and PCa tissue and, more significantly, an increase of 0.1186 AUC for discrimination of insPCa tissue and csPCa tissue. Furthermore, clinical interpretation of CDI^s images by an experienced radiologist shows that CDI^s produces superior contrast between healthy tissue and PCa when compared to T2w and K^{trans}, and provides better contrast between insPCa and csPCa than ADC.

Chapter 4 describes a gradient-based optimization framework through which the parameters of CDI^s may be optimized for particular datasets and clinical tasks. Numerical challenges and parameter constraints were discussed and subsequently addressed through reparameterization of CDI^s and introduction of new hyperparameters, while still retaining its basic definition. Experiments comparing optimized CDI^s to the baseline variants presented in Chapter 3 show that optimization increased AUC for discrimination of healthy tissue and PCa tissue by 0.0105, and improved AUC for discrimination of insPCa tissue and csPCa tissue by 0.0404. Furthermore, comparisons between the MEM and SEM showed

that the MEM often outperforms the SEM for the dataset examined in this thesis, and comparisons of different loss functions showed that a performance trade-off exists between healthy and PCa tissue discrimination and insPCa and csPCa tissue discrimination. Lastly, the impacts of the various hyperparameters on discrimination performance were explored, and the practical considerations of these effects were discussed.

5.2 Clinical Potential

While the numerical results discussed above are compelling from a mathematical standpoint, the ultimate goal of this work is to provide additional value to clinicians beyond that provided by standard MRI modalities. With this in mind, the clinical potential of CDI^s is worth discussing.

First, the radiologist's comparison of CDI^s images and standard MRI images presented in Chapter 3 showed that CDI^s provided better tissue contrast between healthy tissue and insPCa and csPCa tumours. In practice, this contrast may allow radiologists to more easily recognize and prognosticate PCa from MRI imaging alone, which may reduce the need for prostate biopsies (which are uncomfortable for patients and can have negative side effects). Moreover, when biopsies are performed, CDI^s may help to select good biopsy locations and provide confirmation of biopsy results. This is especially valuable when assessing PCa significance, as CDI^s may help to confirm insPCa and identify patients eligible for active surveillance instead of immediate treatment.

Second, the quantitative results presented in Chapter 4 showed that CDI^s achieves significantly higher AUC for discrimination of healthy, insPCa, and csPCa voxels than standard MRI modalities. This suggests that CDI^s better localizes PCa, which may be valuable for guiding prostate biopsies and more accurately and easily segmenting PCa tumours. Considering the increasing use of machine learning in clinical practice, CDI^s may allow for the creation of machine learning models with better diagnostic, prognostic, or segmentation performance compared to models trained with standard MRI images.

5.3 Limitations

While the results presented in this thesis are promising, some limitations remain. First and foremost, it should be noted that although AUC is often correlated with clinical utility, moderate increases in AUC may not correspond to measurable improvements in practice. Moderate differences in AUC may not correspond to visually distinct CDI^s images,

and inter-observer variability may also affect the true clinical efficacy of CDI^s. Simply examining differences in AUC cannot indicate whether CDI^s would improve radiologists' performance in identifying PCa or discriminating insPCa from csPCa, for example. As such, further studies are necessary to establish whether CDI^s has value in practice. Such studies could also include machine learning approaches based on CDI^s (see Section 5.4), as such techniques may better leverage CDI^s in cases where visual analysis does not reveal significant benefits.

The MRI dataset examined in this study was acquired at a single institution. Consistent DWI protocols were used for all patients, and all DWI images were acquired using one of two types of scanners manufactured by the same company. As a result, the experimental results presented in this thesis are not sufficient to show whether CDI^s SI remains consistent across different scanners and imaging protocols. A larger, multi-institutional study would be required to explore this aspect of CDI^s. Moreover, given that only 200 patients imaged in a single country are represented in the dataset, CDI^s has not been evaluated with a diverse cohort in this thesis.

When evaluating the optimization framework in Chapter 4, the same MRI data was used for both training and evaluation of CDI^s. As a result, the reported AUC numbers for optimized AUC may be biased high. Notably, the limited number of parameters in CDI^s likely reduces the risk of such a bias (through a reduced risk of overfitting), but a more rigorous approach would be to retain a dedicated holdout test set using a larger dataset. Ideally, CDI^s could be optimized using data from one institution and then evaluated using data from a separate institution to investigate its generalization properties.

5.4 Future Work

5.4.1 Continuous Synthetic Correlated Diffusion Imaging

While Chapter 4 provided a mechanism by which appropriate b-values could be selected indirectly via optimization, it still considers the choice of b-values to be discrete and parameterized by a particular number and range of b-values. However, DWI decay curves are continuous, and as such a logical extension is to compute CDI^s continuously as well.

Consider a general DW signal decay function $S(b; \boldsymbol{\theta})$, with $\boldsymbol{\theta}$ representing the parameters of the decay model. For example, $\boldsymbol{\theta}$ may represent the parameters of the MEM or SEM. Supposing only synthetic signal acquisitions are used, the signal mixing function M

defined in Equation (3.1) may be expressed as a function of $\mathbf{b}$ with parameters $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$:

$$M_s(\mathbf{b};\boldsymbol{\theta},\boldsymbol{\rho}) = \prod_{i=1}^{N} S(b_i;\boldsymbol{\theta})^{\rho_i} \tag{5.1}$$

This definition may be extended to the continuous domain by combining signals in the form of a product integral. For a given b-value interval $[b_a, b_b]$, the continuous form of Equation (5.1) may be expressed as

$$I(b_a, b_b;\boldsymbol{\theta},\mathbf{w}) = \prod_{b_a}^{b_b} \left(S(b;\boldsymbol{\theta})^{\rho(b;\mathbf{w})}\right)^{db} = \exp\left(\int_{b_a}^{b_b} \rho(b;\mathbf{w}) \ln S(b;\boldsymbol{\theta}) db\right) \tag{5.2}$$

where $\rho : [b_a, b_b] \rightarrow \mathbb{R}$ is a function to control the contributions of different b-values, parameterized by $\mathbf{w}$. Notably, ρ need not be defined exclusively on $[b_a, b_b]$, but its domain must include this interval; for example, $\rho : \mathbb{R} \rightarrow \mathbb{R}$ would also be valid. By replacing the signal mixing portion of Equation (3.5) with Equation (5.2), the continuous form of CDI^s is obtained:

$$C_c(\mathbf{x}; b_a, b_b) = \frac{1}{Z} \iiint_{V(\mathbf{x})} I(b_a, b_b; \hat{\boldsymbol{\theta}}(\mathbf{x}'), \mathbf{w}) f(\mathbf{x}')\, d\mathbf{x}' \tag{5.3}$$

where $\hat{\boldsymbol{\theta}}(\mathbf{x}')$ denotes the estimated parameters of the signal decay model at spatial location $\mathbf{x}'$, and all other symbols are as defined in Equation (3.5). In terms of numerical stability and flexibility, the techniques explored in Section 4.3 (i.e., the geometric mean and gain term) extend naturally to the continuous domain and may therefore be easily adapted for continuous CDI^s.

By using a continuous interval rather than a discrete set of b-values, there is no longer a need to select an appropriate uniform spacing of b-values within the interval. This property may be particularly useful when CDI^s is applied to other diseases, clinical tasks, or datasets, as the optimal spacing of b-values may not generally be consistent in such cases. Moreover, depending on the choice of ρ, this approach may be much more computationally efficient; rather than iteratively computing and multiplying many synthetic DWI acquisitions, the integral in Equation (5.2) may be solved in closed-form and computed non-iteratively.

5.4.2 End-to-end Learning of Predictive Models

In recent years, deep neural networks have been applied to a variety of medical image analysis tasks. These networks typically take medical images as inputs and make predictions based on them. For example, a network may predict whether a patient has PCa,

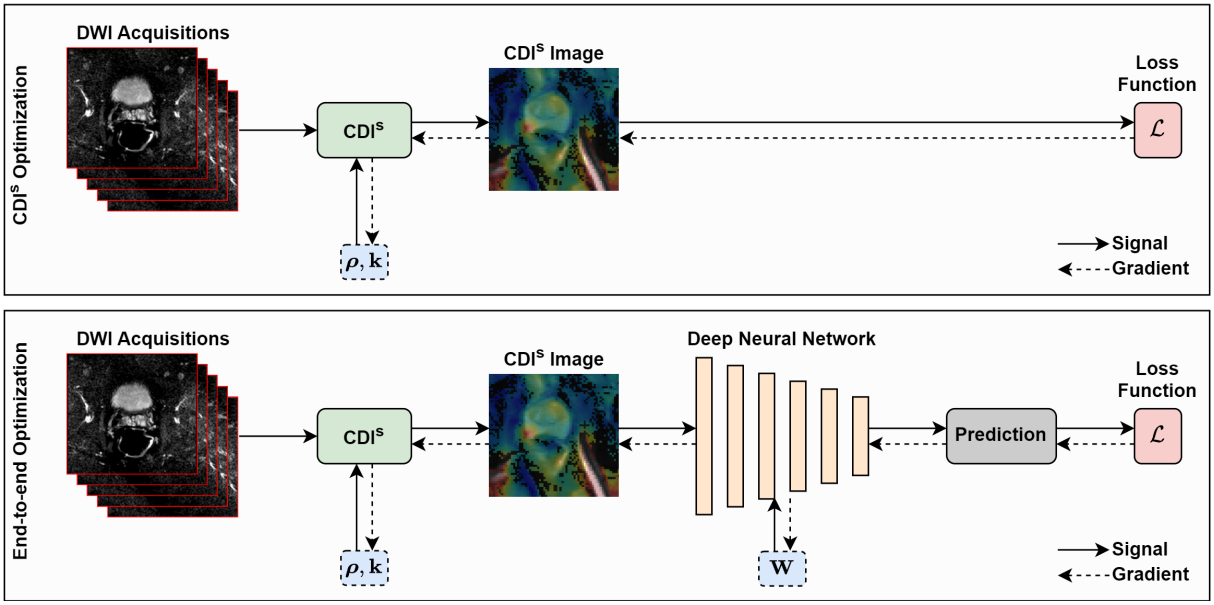

Figure 5.1: Comparison of CDI^s optimization and end-to-end optimization. The top pane illustrates the optimization framework proposed in Chapter 4, where CDI^s images are directly optimized to minimize some loss function. By adding a deep neural network to the framework, the parameters of CDI^s (i.e., $\boldsymbol{\rho}$, $\mathbf{k}$) and the deep network (i.e., $\mathbf{W}$) may be simultaneously optimized for a particular prediction task, as both CDI^s and the network are differentiable functions of their respective parameters.

the GS/GG of a particular tumour, or a mask indicating the boundary of a PCa tumour. However, neural networks are limited by their input data, meaning that poor-quality data can lead to poor predictive performance.

Given that CDI^s may provide better delineation of healthy, insPCa, and csPCa tissues (as explored in this thesis), a logical extension of this thesis would be to train deep neural networks which take CDI^s images as input. Moreover, the optimization framework presented in Chapter 4 presents a unique opportunity to jointly optimize both CDI^s and a deep neural network simultaneously. In doing so, the parameters of CDI^s can be optimized for a specific clinical prediction task rather than for an image-based metric such as AUC, which may improve performance over separate optimization of CDI^s and the network.

Figure 5.1 illustrates how simultaneous optimization could work in practice. The top pane illustrates how CDI^s optimization was performed in this thesis. From DWI inputs, CDI^s images are computed and the parameters of CDI^s are optimized based on a loss func-

tion. However, this loss function may not be directly tied to a clinical task; for example, AUC was optimized (by proxy) in this thesis, but increased AUC may not directly translate to increased clinical utility. By combining CDI^s and a deep neural network sequentially (as in the bottom pane of Figure 5.1), the output becomes a prediction (e.g., PCa presence/absence or PCa grade) which has clear, direct clinical utility, and CDI^s is therefore being optimized directly for a desired clinical use-case.

5.4.3 Application to Other Diseases and Clinical Tasks

An important observation is that the CDI^s optimization framework proposed in this thesis is not specific to PCa. Aside from the particular loss functions, the optimization techniques are solely based on the mathematical properties of DWI, which are generally consistent across many different biological tissues. This presents an opportunity to apply CDI^s and CDI^s optimization whenever DWI is used in clinical practice. For example, DWI is frequently used for imaging of non-prostate cancers, including cancers of the breasts [57], brain [81, 82], and liver [60]. DWI is also commonly used for evaluation of acute stroke [37, 38], and has been studied for assessment of liver fibrosis and cirrhosis [36, 83].

To apply CDI^s to other diseases, an appropriate objective function can be identified which corresponds to the desired clinical task. For example, segmentation of acute stroke lesions allows for measurement of stroke infarct volume, which predicts patient disability and has value in clinical trials [84]. By optimizing CDI^s for delineation of healthy brain tissue and infarct tissue (e.g., using the same soft-AUC loss proposed in this thesis), CDI^s may allow for easier and more accurate manual segmentation. Additionally, by combining CDI^s with a deep segmentation network as described in Section 5.4.2, fully automatic segmentation may be achieved, and CDI^s may improve the performance of such a network beyond that achieved by DWI alone.

References

[1] R. Toivanen and M. M. Shen. Prostate organogenesis: tissue induction, hormonal regulation and cell type specification. *Development*, 144(8):1382–1398, 2017.

[2] L. Zhang, Y. Wang, Z. Qin, et al. Correlation between Prostatitis, Benign Prostatic Hyperplasia and Prostate Cancer: A systematic review and Meta-analysis. *Journal of Cancer*, 11(1):177–189, 2020.

[3] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021. doi: doi.org/10.3322/caac.21660.

[4] Canadian Cancer Statistics Advisory in collaboration with the Canadian Cancer Society and Statistics Canada and the Public Health Agency of Canada. Canadian cancer statistics: A 2022 special report on cancer prevalence. Canadian Cancer Society, 2022. URL http://cancer.ca/Canadian-Cancer-Statistics-2022-EN.

[5] American Cancer Society. Cancer facts & figures 2021. American Cancer Society, 2021. URL https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html.

[6] Ulf-Håkan Stenman, Jari Leinonen, Wan-Ming Zhang, and Patrik Finne. Prostate-specific antigen. *Seminars in Cancer Biology*, 9(2):83–93, 1999. doi: 10.1006/scbi.1998.0086.

[7] R. Chou, J. M. Croswell, T. Dana, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, 155(11):762–771, 2011. doi: 10.7326/0003-4819-155-11-201112060-00375.

[8] S. Loeb, M. A. Bjurlin, J. Nicholson, et al. Overdiagnosis and overtreatment of prostate cancer. *European Urology*, 65(6):1046–1055, 2014. doi: 10.1016/j.eururo.2013.12.062.

[9] L. Klotz, D. Vesprini, P. Sethukavalan, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. *Journal of Clinical Oncology*, 33 (3):272–277, 2015.

[10] C. J. Harvey, J. Pilcher, J. Richenberg, et al. Applications of transrectal ultrasound in prostate cancer. *British Journal of Radiology*, 85(Spec. Iss. 1):3–17, 2012. doi: 10.1259/bjr/56357549.

[11] S. Annunziata, D. A. Pizzuto, and G. Treglia. Diagnostic Performance of PET Imaging Using Different Radiopharmaceuticals in Prostate Cancer According to Published Meta-Analyses. *Cancers (Basel)*, 12(8), 2020. doi: 10.3390/cancers12082153.

[12] S. E. Lapi, H. Wahnishe, D. Pham, et al. Assessment of an 18F-labeled phosphoramidate peptidomimetic as a new prostate-specific membrane antigen-targeted imaging agent for prostate cancer. *Journal of Nuclear Medicine*, 50(12):2042–2048, 2009. doi: 10.2967/jnumed.109.066589.

[13] E. J. Lepin, J. V. Leyton, Y. Zhou, et al. An affinity matured minibody for PET imaging of prostate stem cell antigen (PSCA)-expressing tumors. *European Journal of Nuclear Medicine and Molecular Imaging*, 37(8):1529–1538, 2010. doi: 10.1007/s00259-010-1433-1.

[14] G. Giovacchini, M. Picchio, V. Scattoni, et al. PSA doubling time for prediction of [(11)C]choline PET/CT findings in prostate cancer patients with biochemical failure after radical prostatectomy. *European Journal of Nuclear Medicine and Molecular Imaging*, 37(6):1106–1116, 2010. doi: 10.1007/s00259-010-1403-7.

[15] J. C. Weinreb, J. O. Barentsz, P. L. Choyke, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *European Urology*, 69(1):16–40, 2016.

[16] V. S. Khoo, A. R. Padhani, S. F. Tanner, et al. Comparison of MRI with CT for the radiotherapy planning of prostate cancer: a feasibility study. *British Journal of Radiology*, 72(858):590–597, 1999.

[17] M. Debois, R. Oyen, F. Maes, et al. The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer. *International Journal of Radiation Oncology, Biology, Physics*, 45(4):857–865, 1999.

[18] A. S. Jackson, S. A. Reinsberg, S. A. Sohaib, et al. Distortion-corrected T2 weighted MRI: a novel approach to prostate radiotherapy planning. *British Journal of Radiology*, 80(959):926–933, 2007.

[19] P. Steiger and H. C. Thoeny. Prostate MRI based on PI-RADS version 2: how we review and report. *Cancer Imaging*, 16:9, 2016.

[20] P. Boonsirikamchai, S. Choi, S. J. Frank, et al. MR imaging of prostate cancer in radiation oncology: what radiologists need to know. *RadioGraphics*, 33(3):741–761, 2013.

[21] Y. J. Choi, J. K. Kim, N. Kim, et al. Functional MR imaging of prostate cancer. *RadioGraphics*, 27(1):63–75, 2007.

[22] A. Wong, J. Glaister, A. Cameron, and M. A. Haider. Correlated diffusion imaging. *BMC Medical Imaging*, 13:26, 2013.

[23] F. Khalvati, A. Wong, and M. A. Haider. Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *BMC Medical Imaging*, 15:27, 2015.

[24] B. Delahunt, R. J. Miller, J. R. Srigley, et al. Gleason grading: past, present and future. *Histopathology*, 60(1):75–86, 2012.

[25] G. Ploussard, J. I. Epstein, R. Montironi, et al. The contemporary concept of significant versus insignificant prostate cancer. *European Urology*, 60(2):291–303, 2011.

[26] J. I. Epstein, L. Egevad, M. B. Amin, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *American Journal of Surgical Pathology*, 40(2):244–252, 2016.

[27] P. A. Humphrey, H. Moch, A. L. Cubilla, et al. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours. *European Urology*, 70(1):106–119, 2016.

[28] G. J. L. H. van Leenders, T. H. van der Kwast, D. J. Grignon, et al. The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. *American Journal of Surgical Pathology*, 44(8):e87–e99, 2020.

[29] J. L. Mohler, E. S. Antonarakis, A. J. Armstrong, et al. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network*, 17(5):479–505, 2019.

[30] L. S. Mertens, P. J. van Leeuwen, and H. G. van der Poel. Abandon the Label of Clinically Insignificant Prostate Cancer. *European Urology Open Science*, 37:36–37, 2022.

[31] J. J. Tosoian, M. Mamawala, J. I. Epstein, et al. Active Surveillance of Grade Group 1 Prostate Cancer: Long-term Outcomes from a Large Prospective Cohort. *European Urology*, 77(6):675–682, 2020.

[32] S. P. Porten, J. M. Whitson, J. E. Cowan, et al. Changes in prostate cancer grade on serial biopsy in men undergoing active surveillance. *Journal of Clinical Oncology*, 29 (20):2795–2800, 2011.

[33] J. K. Mullins, D. Bonekamp, P. Landis, et al. Multiparametric magnetic resonance imaging findings in men with low-risk prostate cancer followed using active surveillance. *BJU International*, 111(7):1037–1045, 2013.

[34] M. H. Maurer and J. T. Heverhagen. Diffusion weighted imaging of the prostate-principles, application, and advances. *Translational Andrology and Urology*, 6(3): 490–498, 2017.

[35] S. C. Partridge and E. S. McDonald. Diffusion weighted magnetic resonance imaging of the breast: protocol optimization, interpretation, and clinical applications. *Magnetic Resonance Imaging Clinics of North America*, 21(3):601–624, 2013.

[36] M. K. Shin, J. S. Song, S. B. Hwang, et al. Liver Fibrosis Assessment with Diffusion-Weighted Imaging: Value of Liver Apparent Diffusion Coefficient Normalization Using the Spleen as a Reference Organ. *Diagnostics (Basel)*, 9(3), 2019.

[37] K. J. van Everdingen, J. van der Grond, L. J. Kappelle, et al. Diffusion-weighted magnetic resonance imaging in acute stroke. *Stroke*, 29(9):1783–1790, 1998. doi: 10.1161/01.STR.29.9.1783. URL https://www.ahajournals.org/doi/abs/10.1161/01.STR.29.9.1783.

[38] G. W. Albers. Diffusion-weighted MRI for evaluation of acute stroke. *Neurology*, 51 (3 Suppl 3):S47–49, 1998.

[39] M. H. Maurer and J. T. Heverhagen. Diffusion weighted imaging of the prostate-principles, application, and advances. *Transl Androl Urol*, 6(3):490–498, Jun 2017.

[40] V. P. Grover, J. M. Tognarelli, M. M. Crossey, et al. Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. *Journal of Clinical and Experimental Hepatology*, 5(3):246–255, 2015.

[41] G. S. Chilla, C. H. Tan, C. Xu, and C. L. Poh. Diffusion weighted magnetic resonance imaging and its recent trend-a survey. *Quantitative Imaging in Medicine and Surgery*, 5(3):407–422, 2015.

[42] Idoia Corcuera-Solano, Mathilde Wagner, Stefanie Hectors, et al. DWI of the prostate: Comparison of a faster diagonal acquisition to standard three-scan trace acquisition. *Journal of Magnetic Resonance Imaging*, 46(6):1767–1775, 2017.

[43] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics*, 42: 288–292, 1965. doi: 10.1063/1.1695690.

[44] V. Brancato, C. Cavaliere, M. Salvatore, and S. Monti. Non-Gaussian models of diffusion weighted imaging for detection and characterization of prostate cancer: a systematic review and meta-analysis. *Scientific Reports*, 9(1):16837, 2019.

[45] X. Wang, N. Tu, T. Qin, et al. Diffusion Kurtosis Imaging Combined With DWI at 3-T MRI for Detection and Assessment of Aggressiveness of Prostate Cancer. *American Journal of Roentgenology*, 211(4):797–804, 2018.

[46] N. He, Z. Li, X. Li, et al. Intravoxel Incoherent Motion Diffusion-Weighted Imaging Used to Detect Prostate Cancer and Stratify Tumor Grade: A Meta-Analysis. *Frontiers in Oncology*, 10:1623, 2020.

[47] X. Liu, W. Peng, L. Zhou, and H. Wang. Biexponential apparent diffusion coefficients values in the prostate: comparison among normal tissue, prostate cancer, benign prostatic hyperplasia and prostatitis. *Korean Journal of Radiology*, 14(2):222–232, 2013.

[48] X. Liu, L. Zhou, W. Peng, et al. Comparison of stretched-Exponential and mono-exponential model diffusion-Weighted imaging in prostate cancer and normal tissues. *Journal of Magnetic Resonance Imaging*, 42(4):1078–1085, 2015.

[49] Jussi Toivonen, Harri Merisaari, Marko Pesola, et al. Mathematical models for diffusion-weighted imaging of prostate cancer using b values up to 2000 s/mm2: Correlation with gleason score and repeatability of region of interest analysis. *Magnetic Resonance in Medicine*, 74(4):1116–1124, 2015. doi: 10.1002/mrm.25482.

[50] H. J. Meyer, A. Wienke, and A. Surov. Discrimination between clinical significant and insignificant prostate cancer with apparent diffusion coefficient - a systematic review and meta analysis. *BMC Cancer*, 20(1):482, 2020.

[51] J. H. Yim, C. K. Kim, and J. H. Kim. Clinically insignificant prostate cancer suitable for active surveillance according to Prostate Cancer Research International: Active surveillance criteria: Utility of PI-RADS v2. *Journal of Magnetic Resonance Imaging*, 47(4):1072–1079, 2018.

[52] K. M. Bennett, K. M. Schmainda, R. Bennett (Tong), et al. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magnetic Resonance in Medicine*, 50(4):727–734, 2003. doi: https://doi.org/10.1002/mrm.10581.

[53] E. Kim, C. K. Kim, H. S. Kim, et al. Histogram analysis from stretched exponential model on diffusion-weighted imaging: evaluation of clinically significant prostate cancer. *British Journal of Radiology*, 93(1106):20190757, 2020.

[54] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944. doi: 10.1090/qam/10666.

[55] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. doi: 10.1137/0111030.

[56] H. Li, L. Liang, A. Li, et al. Monoexponential, biexponential, and stretched exponential diffusion-weighted imaging models: Quantitative biomarkers for differentiating renal clear cell carcinoma and minimal fat angiomyolipoma. *Journal of Magnetic Resonance Imaging*, 46(1):240–247, 2017.

[57] Z. M. Almutlaq, D. J. Wilson, S. E. Bacon, et al. Evaluation of Monoexponential, Stretched-Exponential and Intravoxel Incoherent Motion MRI Diffusion Models in Early Response Monitoring to Neoadjuvant Chemotherapy in Patients With Breast Cancer-A Preliminary Study. *Journal of Magnetic Resonance Imaging*, 56(4):1079–1088, 2022.

[58] Z. Feng, X. Min, D. J. A. Margolis, et al. Evaluation of different mathematical models and different b-value ranges of diffusion-weighted imaging in peripheral zone prostate cancer detection using b-value up to 4500 s/mm^2. *PLoS ONE*, 12(2):e0172127, 2017. doi: 10.1371/journal.pone.0172127.

[59] Y. Mazaheri, A. M. Hötker, A. Shukla-Dave, et al. Model selection for high b-value diffusion-weighted MRI of the prostate. *Magnetic Resonance Imaging*, 46:21–27, 2018.

[60] Y. Zhou, H. X. Zhang, X. S. Zhang, et al. Non-mono-exponential diffusion models for assessing early response of liver metastases to chemotherapy in colorectal cancer. *Cancer Imaging*, 19(1):39, 2019.

[61] X. Min, Z. Feng, L. Wang, et al. Multi-model analysis of diffusion-weighted imaging of normal testes at 3.0 T: Preliminary findings. *Academic Radiology*, 25(4):445–452, 2018.

[62] H. Merisaari, H. Laakso, H. Liljenbäck, et al. Statistical Evaluation of Different Mathematical Models for Diffusion Weighted Imaging of Prostate Cancer Xenografts in Mice. *Frontiers in Oncology*, 11:583921, 2021.

[63] S. J. Hectors, S. Semaan, C. Song, et al. Advanced Diffusion-weighted Imaging Modeling for Prostate Cancer Characterization: Correlation with Quantitative Histopathologic Tumor Tissue Composition-A Hypothesis-generating Study. *Radiology*, 286(3): 918–928, 2018.

[64] Y. R. Ueno, T. Tamada, S. Takahashi, et al. Computed Diffusion-Weighted Imaging in Prostate Cancer: Basics, Advantages, Cautions, and Future Prospects. *Korean Journal of Radiology*, 19(5):832–837, 2018.

[65] S. Jendoubi, M. Wagner, S. Montagne, et al. MRI for prostate cancer: can computed high b-value DWI replace native acquisitions? *European Radiology*, 29(10):5197–5204, 2019.

[66] P. Ning, D. Shi, G. A. Sonn, et al. The impact of computed high b-value images on the diagnostic accuracy of DWI for prostate cancer: A receiver operating characteristics analysis. *Scientific Reports*, 8(1):3409, 2018.

[67] A. B. Rosenkrantz, H. Chandarana, N. Hindman, et al. Computed diffusion-weighted imaging of the prostate at 3 T: impact on image quality and tumour detection. *European Radiology*, 23(11):3170–3177, 2013.

[68] A. B. Rosenkrantz, N. Parikh, A. S. Kierans, et al. Prostate Cancer Detection Using Computed Very High b-value Diffusion-weighted Imaging: How High Should We Go? *Academic Radiology*, 23(6):704–711, 2016.

[69] H. K. Agarwal, F. V. Mertan, S. Sankineni, et al. Optimal high b-value for diffusion weighted MRI in diagnosing high risk prostate cancers in the peripheral zone. *Journal of Magnetic Resonance Imaging*, 45(1):125–131, 2017.

[70] Geert Litjens, Oscar Debats, Jelle Barentsz, et al. Computer-aided detection of prostate cancer in mri. *IEEE Transactions on Medical Imaging*, 33(5):1083–1092, 2014. doi: 10.1109/TMI.2014.2303821.

[71] Geert Litjens, Oscar Debats, Jelle Barentsz, et al. Prostatex challenge data [data set]. The Cancer Imaging Archive, 2017.

[72] K. Clark, B. Vendt, K. Smith, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.

[73] Renato Cuocolo, Arnaldo Stanzione, Anna Castaldo, et al. Quality control and whole-gland, zonal and lesion annotations for the prostatex challenge public dataset. *European Journal of Radiology*, 138:109647, 2021. doi: 10.1016/j.ejrad.2021.109647.

[74] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3):837–845, 1988.

[75] J. M. Bland and D. G. Altman. Multiple significance tests: the Bonferroni method. *BMJ*, 310(6973):170, 1995.

[76] Adam Paszke, Sam Gross, Francisco Massa, et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[77] Toon Calders and Szymon Jaroszewicz. Efficient auc optimization for classification. In Joost N. Kok, Jacek Koronacki, Ramon Lopez de Mantaras, et al., editors, *Knowledge Discovery in Databases: PKDD 2007*, pages 42–53, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-74976-9.

[78] Lian Yan, Robert Dodier, Michael C. Mozer, and Richard Wolniewicz. Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic.

In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ICML'03, page 848–855. AAAI Press, 2003. ISBN 1577351894.

[79] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151, 1999.

[80] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations*, 2017.

[81] X. Chen, J. Jiang, N. Shen, et al. Stretched-exponential model diffusion-weighted imaging as a potential imaging marker in preoperative grading and assessment of proliferative activity of gliomas. *American Journal of Translational Research*, 10(8): 2659–2668, 2018.

[82] M. Kusunoki, K. Kikuchi, O. Togao, et al. Differentiation of high-grade from low-grade diffuse gliomas using diffusion-weighted imaging: a comparative study of mono-, bi-, and stretched-exponential diffusion models. *Neuroradiology*, 62(7):815–823, 2020.

[83] K. Sandrasegaran, F. M. Akisik, C. Lin, et al. Value of diffusion-weighted MRI for assessing liver fibrosis and cirrhosis. *AJR Am J Roentgenol*, 193(6):1556–1560, 2009.

[84] K. K. Wong, J. S. Cummock, G. Li, et al. Automatic Segmentation in Acute Ischemic Stroke: Prognostic Significance of Topological Stroke Volumes on Stroke Outcome. *Stroke*, 53(9):2896–2905, 2022.

APPENDICES

Appendix A

Derivations

A.1 Invariance to Intensity Scaling

Suppose N DWI acquisitions (either native or synthetic) are used to compute CDI^s SI without calibration factor Z:

$$C_u(\mathbf{x};\mathbf{b},\boldsymbol{\rho}) = Z \cdot C_s(\mathbf{x};\mathbf{b},\boldsymbol{\rho}) = \iiint_{V(\mathbf{x})} \prod_{i=1}^{N} \left(\tilde{S}_{b_i}(\mathbf{x})^{\rho_i}\right) f\left(\mathbf{x}';\tilde{S}_{b_1}(\mathbf{x}'),\ldots,\tilde{S}_{b_N}(\mathbf{x}')\right) d\mathbf{x}'$$

where f is as defined in Equation (3.5), but is written differently to stress that it may depend on DWI SI. Now, suppose the DWI acquisitions are multiplied by a constant scalar

$a \neq 0$, yielding $\tilde{S}'_{b_i} = a\tilde{S}_{b_i}$. Then, the uncalibrated SI from the scaled DWI acquisitions is

$$
\begin{aligned}
C'_u(\mathbf{x};\mathbf{b},\boldsymbol{\rho}) &= \iiint\limits_{V(\mathbf{x})} \prod_{i=1}^{N} \left(\tilde{S}'_{b_i}(\mathbf{x})^{\rho_i}\right) f\left(\mathbf{x}';\tilde{S}'_{b_1}(\mathbf{x}'),\ldots,\tilde{S}'_{b_N}(\mathbf{x}')\right) d\mathbf{x}' && \text{(A.1)}\\
&= \iiint\limits_{V(\mathbf{x})} \prod_{i=1}^{N} \left(\left(a\tilde{S}_{b_i}(\mathbf{x})\right)^{\rho_i} \right) f\left(\mathbf{x}';a\tilde{S}_{b_1}(\mathbf{x}'),\ldots,a\tilde{S}_{b_N}(\mathbf{x}')\right) d\mathbf{x}' && \text{(by def. of } \tilde{S}'_{b_i})\\
&= \iiint\limits_{V(\mathbf{x})} \prod_{i=1}^{N} \left(\left(a\tilde{S}_{b_i}(\mathbf{x})\right)^{\rho_i} \right) f\left(\mathbf{x}';\tilde{S}_{b_1}(\mathbf{x}'),\ldots,\tilde{S}_{b_N}(\mathbf{x}')\right) d\mathbf{x}' && \text{(by invariance of } f)\\
&= a^{\sum_{i=1}^{N}\rho_i} \iiint\limits_{V(\mathbf{x})} \prod_{i=1}^{N} \left(\tilde{S}_{b_i}(\mathbf{x})^{\rho_i}\right) f\left(\mathbf{x}';\tilde{S}_{b_1}(\mathbf{x}'),\ldots,\tilde{S}_{b_N}(\mathbf{x}')\right) d\mathbf{x}' && \text{(by linearity)}\\
&= a^{\sum_{i=1}^{N}\rho_i} C_u(\mathbf{x};\mathbf{b},\boldsymbol{\rho})
\end{aligned}
$$

Since $a^{\sum_{i=1}^{N}\rho_i}$ is constant and the median function is homogeneous of degree 1, the median obtained from $C'_u(\mathbf{x};\mathbf{b},\boldsymbol{\rho})$ is simply

$$Z' = a^{\sum_{i=1}^{N}\rho_i} Z \tag{A.2}$$

Thus, the final SI obtained from the scaled DWI acquisitions, after calibration, is

$$
\begin{aligned}
C'_s(\mathbf{x};\mathbf{b},\boldsymbol{\rho}) &= \frac{1}{Z'} C'_u(\mathbf{x};\mathbf{b},\boldsymbol{\rho})\\
&= \frac{a^{\sum_{i=1}^{N}\rho_i}}{a^{\sum_{i=1}^{N}\rho_i} Z} C_u(\mathbf{x};\mathbf{b},\boldsymbol{\rho})\\
&= \frac{1}{Z} C_u(\mathbf{x};\mathbf{b},\boldsymbol{\rho})\\
&= C_s(\mathbf{x};\mathbf{b},\boldsymbol{\rho})
\end{aligned} \tag{A.3}
$$

and therefore constant non-zero scaling of DWI SI is corrected by CDI^s calibration.

A.2 Soft-AUC Loss

The soft-AUC loss is defined as the complement of soft-AUC (Equation (4.18)):

$$\begin{aligned}
\mathcal{L}(\mathcal{D}_n, \mathcal{D}_p; \beta) &= 1 - \mathrm{AUC}_{\mathrm{soft}}(\mathcal{D}_n, \mathcal{D}_p; \beta) \\
&= 1 - \frac{\sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} \sigma_\beta(p - n)}{|\mathcal{D}_n||\mathcal{D}_p|} \\
&= \frac{|\mathcal{D}_n||\mathcal{D}_p| \sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} \sigma_\beta(p - n)}{|\mathcal{D}_n||\mathcal{D}_p|} \\
&= \frac{\sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} (1 - \sigma_\beta(p - n))}{|\mathcal{D}_n||\mathcal{D}_p|} \\
&= \frac{\sum_{n \in \mathcal{D}_n} \sum_{p \in \mathcal{D}_p} \sigma_\beta(n - p)}{|\mathcal{D}_n||\mathcal{D}_p|} \qquad (\text{since } 1 - \sigma_\beta(x) = \sigma_\beta(-x))
\end{aligned} \tag{A.4}$$