QUANTITATIVE MAGNETIC RESONANCE IMAGING AND ANALYSIS OF ARTICULAR CARTILAGE AND OSTEOARTHRITIS

Dimitri Alexander Kessler

Fitzwilliam College

This dissertation is submitted for the degree of

Doctor of Philosophy

Department of Radiology

University of Cambridge

September 2021

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Degree Committee for Clinical Medicine & Veterinary Medicine.

Dimitri Alexander Kessler
Cambridge, September 2021

Summary

Quantitative Magnetic Resonance Imaging and Analysis of Articular Cartilage and Osteoarthritis

Dimitri Alexander Kessler

MRI plays an important role in the continuing search for a sensitive osteoarthritis (OA) imaging biomarker able to detect early, pre-morphological alterations in cartilage composition. Determining the compositional recovery pattern of cartilage following acute joint loading could potentially present a more sensitive biomarker for defining cartilage health [1]. However, only a limited amount of studies have assessed both the immediate effect of joint loading on cartilage, as well as its post-loading recovery. In addition, when assessing the compositional responses of cartilage to joint loading, previous studies usually did not incorporate the measurement error of the used quantitative MRI technique into their analysis. Therefore, an uncertainty persists whether or not compositional MRI techniques are sensitive enough to measure changes in water and macromolecular content of cartilage, or if previous studies were merely measuring noise. Consequently, an objective of this thesis is to increase our understanding of and reliability in quantitative T_2 and $T_{1\rho}$ relaxation time mapping to detect compositional responses of cartilage following a joint loading activity.

Furthermore, to obtain the quantitative morphological and compositional measures of cartilage, detailed region-specific delineation of cartilage is required. This delineation (or segmentation) of cartilage is laborious and time-consuming as it is usually performed manually by an expert observer. Many new advances in image analysis, particularly those in convolutional neural networks (CNNs) and deep learning, have enabled a time-efficient semi- or fully-automated alternative to this process [2, 3]. This thesis explores the utility of deep CNNs generated segmentations for accurate surface-based analysis of cartilage morphology and composition from knee MRIs as well as of cortical bone thickness from knee CTs.

Chapter 1 will provide an introduction into the structure and biomechanics of articular cartilage and the role of MRI in imaging the degenerative joint disorder, osteoarthritis as well as the effects of different joint loading activities on cartilage morphology and composition.

Chapter 2 explains the principle of MRI and the pulse sequences used in the following chapter for the morphometric and compositional assessment of articular cartilage.

Chapter 3 describes the use of 3D Cartilage Surface Mapping (3D-CaSM) [3] to assess variations in cartilage $T_{1\rho}$ and T_2 relaxation times of young, healthy participants following a mild, unilateral stepping activity. By evaluating and incorporating the intrasessional repeatability of the $T_{1\rho}$ and T_2 mapping techniques, I aim to highlight those cartilage areas experiencing exercise-induced compositional changes greater than measurement error.

A significant amount of time is needed to manually segment the regions-of-interest required

to perform the 3D-CaSM used in Chapter 3. Therefore, in Chapter 4, I assessed the use of deep convolutional neural networks for automating the segmentation process for multiple knee joint tissues simultaneous and increase the time-efficiency for evaluating knee MR datasets. I evaluated the use of a conditional Generative Adversarial Network (cGAN) as a potentially improved method for automated segmentation compared to the widely used convolutional neural network, U-Net.

In Chapter 5 I combined the 3D-CaSM and automated segmentation methods presented in Chapters 3 and 4, respectively to assess the use of fully automatic segmentations of femoral and tibial bone-cartilage structures for accurate surface-based analysis of cartilage morphology and composition on knee MR images. This was performed on publicly available data from the Osteoarthritis Initiative, a multicentre observational study with expert manual segmentations provided by the Zuse Institute in Berlin.

Chapter 6 describes an automated pipeline for subchondral cortical bone thickness mapping from knee CT data. I developed a method of using automated segmentations of articular cartilage and bone from knee MRI data to determine the periarticular bone surface which is covered by cartilage. This surface was then used to perform cortical bone thickness measurements on corresponding CT data. I validated this pipeline using data from the EU-funded, multi-centre observational study called Applied Private-Public partneRship enabling OsteoArthritis Clinical Headway (APPROACH).

Chapter 7 summarises the main conclusions and contributions of the works presented in this thesis as well as providing directions for future work.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my PhD advisor, Professor Fiona Gilbert, for all her guidance and support over the course of my PhD degree. Her mentorship has been an invaluable experience that I will carry with me throughout my career. Thank you for allowing me to be progressively independent as my PhD progressed, and always providing advice when things got challenging. The co-supervision of Dr Joshua Kaggie has been irreplaceable, especially his expertise in MRI, image analysis and programming. I gratefully thank him for all his time, support and encouragement he has provided throughout the past years and it has been a real pleasure to work with and learn from him.

I am indebted to Dr Jamie MacKay for helping me getting started when I arrived and always being available to discuss new ideas and projects, even after he embarked upon new professional journeys. I am very grateful for all his clinical insights and assistance with image analysis and statistics. I would like to sincerely thank Professor Martin Graves for not only being the most knowledgeable MR expert I have ever met and had the privilege of working with, but also always making the time for anyone in his already overflowing outlook calendars to answer any question, proofread manuscripts or offer advice on professional and personal life decisions.

To the orthopaedic department led by Professor Andrew McCaskie. They have provided a warm welcome and continual support throughout the past years. Thank you to all for their time and resources. Special thanks must go to Jennifer O'Callaghan, Stephen McDonnell and Wasim Khan for their commitment, energy and backing of the FERARI study.

Thank you to all the past and present members of the radiology department, for creating a fun and engaging learning environment, flowing with new and exciting ideas. Thanks for always being open to sharing your research work and ideas at various seminars. I would like to thank Dr Andrew Grainger for his expertise on all things musculoskeletal imaging. A huge thank you to the radiology research nurses, in particular, Amy Frary, for their help with patient recruitment and keeping study documentation on point. A special thank you goes to all previous and current members of the 'coolest office in town' in room 253. I appreciate the time given by Dr Roie Manavaki and Dr Eva Serraro for not only entertaining all my questions (mainly concerning R&D and study amendments) but also providing valuable advice and guidance. The wonderful and loving care of the departmental administrative staff cannot go unmentioned. Specifically, the organisational support and patience shown by Candice Anderson and Sarah Perkins has been unbelievable.

This work, and with it my PhD, could not have been possible without the funding and expert support from GlaxoSmithKline. I have learnt a lot through this academic-industrial collaboration and appreciate the knowledge and advice provided by Rob Janiczek and Alex Morgan-Roberts throughout the set-up and coordination of a prospective clinical imaging trial.

With that in mind, I am unbelievably appreciative to all the people that have participated in our imaging studies. They voluntarily gave up their time to be imaged for up to 3 hours at a time and allow us to perform hopefully impactful research. It has been hugely rewarding to meet all the participants of the prospective clinical OA trial (FERARI) and learn of their struggles living with OA. You have instilled an ongoing motivation in me to continue the search for an effective OA treatment.

A big thank you to my friends in and outside of the department. To Dr Gabrielle Baxter

and Dr Sarah Hickman, thanks for the jokes and enjoyable 'soup times' that helped calm down the nerves at times when obstacles seemed unconquerable. Thank you also to my friends and teammates at the Cambridge Floorball Club for giving me a place to not only welcomingly settle in when I newly arrived in the UK but also where I could unwind and gain new energy during my PhD.

I am very grateful to my parents, Silvia and Frank, and my sister, Isabella, for supporting and encouraging all my life decisions, even if they meant regularly moving away. Starting at a very early age in South Africa and Namibia, my parents have always provided me with an excellent education. I probably would not have had the opportunity to embark on a PhD in Cambridge if I did not have their continuous support and praise.

To Cecilia, my source of inspiration and reassurance, my link in achieving all successes and overcoming all obstacles, my best friend and partner in life. Her love and support have and always will push me forward and accomplish great things, and I cannot wait to spend the rest of my life with her.

Publications and Conference Proceedings

Publications arising from this thesis

Kessler DA, MacKay JW, Crowe V, Henson F, Graves MJ, Gilbert FJ and Kaggie JD. The Optimisation of Deep Neural Networks for Segmenting Multiple Knee Joint Tissues from MRIs. *Computerized Medical Imaging and Graphics* 86, December 2020. https://doi.org/10.1016/j.compmedimag.2020.101793.

Kessler DA, MacKay JW, McDonald S, McDonnell SM, Grainger A, Roberts AR, Janiczek RL, Graves MJ, Kaggie JD and Gilbert FJ. Effectively Measuring Exercise-related Variations in T1ρ and T2 Relaxation Times of Healthy Articular Cartilage. *Journal of Magnetic Resonance Imaging* 52(6), 1753 - 1764, 2020. https://doi.org/10.1002/jmri.27278.

Presentations arising from this thesis

Oral

Kessler DA, Kaggie JD, MacKay JW, Morgan-Roberts AR, Janiczek R, Graves MJ and Gilbert FJ. Quantitative MR Relaxation Imaging of Cartilage Compositional Response to Exercise. In: *Proceedings of the 27th Annual Meeting of ISMRM, Montreal, Canada*, 2019; p 0416.

Poster

Kessler DA, McLean MA, Lanz T, Riemer F, Schulte RF, Grainger A, Gilbert FJ, Graves MJ and Kaggie JD. Bilateral Sodium Magnetic Resonance Imaging of the Lower Extremity. In: *Proceedings of the 28th Annual Meeting of ISMRM*, 2020; p 6277.

Kessler DA, Kaggie JD, MacKay JW, McDonald S, Grainger A, Roberts AR, Janiczek RL, Graves MJ and Gilbert FJ. Distinguishing Exercise-Induced Compositional Changes in Knee Cartilage with Quantitative MR Relaxation Time Mapping. In: *Proceedings of the 28th Annual Meeting of ISMRM*, 2020; p 4548.

Kessler DA, MacKay JW, Gilbert FJ, Graves MJ and Kaggie JD. Automated Segmentation of Knee Articular Cartilage on MRI Data: Increasing Network Capacity with Transfer Learning. In: *Proceedings of the 28th Annual Meeting of ISMRM*, 2020; p 451.

Kessler DA, Gilbert FJ, MacKay JW, Graves MJ and Kaggie JD. Automated Knee MRI Semantic Segmentation with Generative Adversarial Networks. In: *Proceedings of the 27th Annual Meeting of ISMRM*, Montreal, Canada, 2019; p 4808.

Kessler DA, Kaggie JD, MacKay JW, Morgan AR, Janiczek R, Graves MJ and Gilbert FJ. Imaging Compositional Cartilage Response After Exercise Using T1rho and T2 Relaxation Mapping. In: *Proceedings of the 24th Annual Meeting of British Chapter ISMRM*, Oxford, United Kingdom, 2018; p 97 (Poster 09).

Publications arising from work unrelated to this thesis

Ranmuthu CDS., MacKay JW, Crowe VA, Kaggie JD, **Kessler DA** and McDonnell SM. Quantitative analysis of the ACL and PCL using T1rho and T2 relaxation time mapping: an exploratory, cross-sectional comparison between OA and healthy control knees. BMC Musculoskeletal Disorder 22, 916 2021.
https://doi.org/10.1186/s12891-021-04755-y

Kaggie JD, Lanz T, McLean MA, Riemer F, Schulte RF, Benjamin AJV, **Kessler DA**, Sun C, Gilbert FJ, Graves MJ and Gallagher FA. Combined 23Na and 13C imaging at 3.0 Tesla using a single-tuned large FOV birdcage coil. *Magnetic Resonance in Medicine*, 2021.
https://doi.org/10.1002/mrm.28772.

Serrao EM, **Kessler DA**, Carmo B, Beer L, Brindle KM, Buonincontri G, Gallagher FA, Gilbert FJ, Godfrey E, Graves MJ, McLean MA, Sala E, Schulte RF and Kaggie JD. Magnetic resonance fingerprinting of the pancreas at 1.5 T and 3.0 T. *Scientific Reports* 10(1), 17563, 2020.
https://doi.org/10.1038/s41598-020-74462-6.

Kessler DA, MacKay JW, McDonnell S and Kaggie JD. Editorial for "Diffusion Tensor Imaging for Quantitative Assessment of Anterior Cruciate Ligament Injury Grades and Graft". *Journal of Magnetic Resonance Imaging* 52(5), 1485-1486, 2020.
https://doi.org/10.1002/jmri.27317.

Kaggie JD, Deen S, **Kessler DA**, McLean MA, Buonincontri G, Schulte RF, Addley H, Sala E, Brenton J, Graves MJ and Gallagher FA. Feasibility of Quantitative Magnetic Resonance Fingerprinting in Ovarian Tumours for T1 and T2 Mapping in a PET/MR Setting. *IEEE Transactions on Radiation and Plasma Medical Sciences* 3(4), 509-515, 2019.
https://doi.org/10.1109/TRPMS.2019.2905366.

Commonly used Abbreviations and Mathematical Symbols

Abbreviations / Acronyms

2D	Two-dimensional
3D	Three-dimensional
3D-CaSM	Three-dimensional cartilage surface mapping
APPROACH	Applied Public-Private Research enabling OsteoArthritis Clinical Headway
cGAN	Conditional generative adversarial network
CNN	Convolutional neural network
CV	Coefficient of variation
DL	Deep learning
FERARI	Functional Excercise Response on osteoArthritis Imaging
FoV	Field of view
FSE	Fast spin echo
GAN	Generative adversarial network
GRE	Gradient echo
MRI	Magnetic resonance imaging
OA	Osteoarthritis
OAI	Osteoarthritis Initiative
PD	Proton density
RF	Radiofrequency
RMS	Root-mean-squared
SE	Spin echo
SDD	Smallest detectable difference
SKI10	Segmentation of Knee Images 2010
SPGR	Spoiled gradient recalled-echo
$\mathbf{T}_1$	Spin-lattice relaxation time
$\mathbf{T}_{1\rho}$	Spin-lattice relaxation time in the rotation frame
$\mathbf{T}_2$	Spin-spin relaxation time
$\mathbf{T}_2^*$	"Effective" T_2
TE	Echo time
TR	Repetition time

TSL	Spin-lock time
ZIB	Zuse Institute Berlin

Mathematical Symbols

α	Flip angle
γ	Gyromagnetic ratio
$\hbar$	Reduced Planck's constant
μ	Magnetic dipole moment
ω_L	Larmor frequency
$\mathbf{B}_0$	Static magnetic field
$\mathbf{B}_1$	Radiofrequency field
$\mathbf{M}_0$	Net longitudinal magnetisation

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Knee Articular Cartilage and Osteoarthritis

1.1.1 Cartilage Structure, Metabolism and Biomechanics

Hyaline (articular) cartilage is an elastic tissue that covers the synovial, or diarthrodial, joint areas as the articular surface to reduce friction between opposing bone surfaces and to withstand and distribute pressure in the joint. This cartilage primarily consists of an extracellular matrix (ECM) and a small number of sparsely distributed chondrocyte cells. Articular cartilage lacks blood vessels and nerves with mass exchange taking place through diffusion [4]. The most abundant component of articular cartilage and the ECM is water, with other lesser concentrated components being proteoglycans and multiple types of collagen fibres. The cellular organisation and collagen fibre architecture of articular cartilage can be divided into four zones - the thin superficial zone (10%-20% of total thickness), the middle zone (40%-60% of total thickness), the deep zone (30% of total thickness) and the calcified zone [5–7]. The tidemark characterises the interface between the deep zone and the calcified cartilage layer [6]

Collagen fibrils within articular cartilage are primarily type II collagen with the lesser populated collagen types (such as XI and IX) being there to stabilise the fibrils structures [7]. When expanded, collagen fibrils are stiff and strong, however offer minimal resistance when compressed due to their large length-to-thickness ratio. In the thin superficial zone, the collagen fibrils are densely packed and orientated parallel to the cartilage surface to protect and maintain the deeper cartilage layers. The fibrils in the middle zone are less compressed and randomly orientated, contributing to the resistant response of cartilage to shear stress [7]. In the deep zone, the collagen fibrils compose larger, parallel aligned bundles that are fixed perpendicular through the calcified cartilage layer to the subchondral bone [6, 7]. This variation in orientation, size and density of the collagen fibrils is fundamental to the ability of cartilage to withstand mechanical pressure [6, 7].

Articular cartilage contains large aggregating proteoglycans (aggrecans) along with several smaller proteoglycans. Proteoglycans consist of a protein core with one or more (up to few hundred) negatively charged glycosaminoglycans (GAGs) attached to it. This negative charge within the cartilage tissue attracts counterions, such as sodium ions, which in turn results in an uptake of water by osmotic processes. Thus, a crucial function of the proteoglycans is to attract water and maintain turgor to enable the cartilage tissue to cope with pressure and distribute mechanical stress [8].

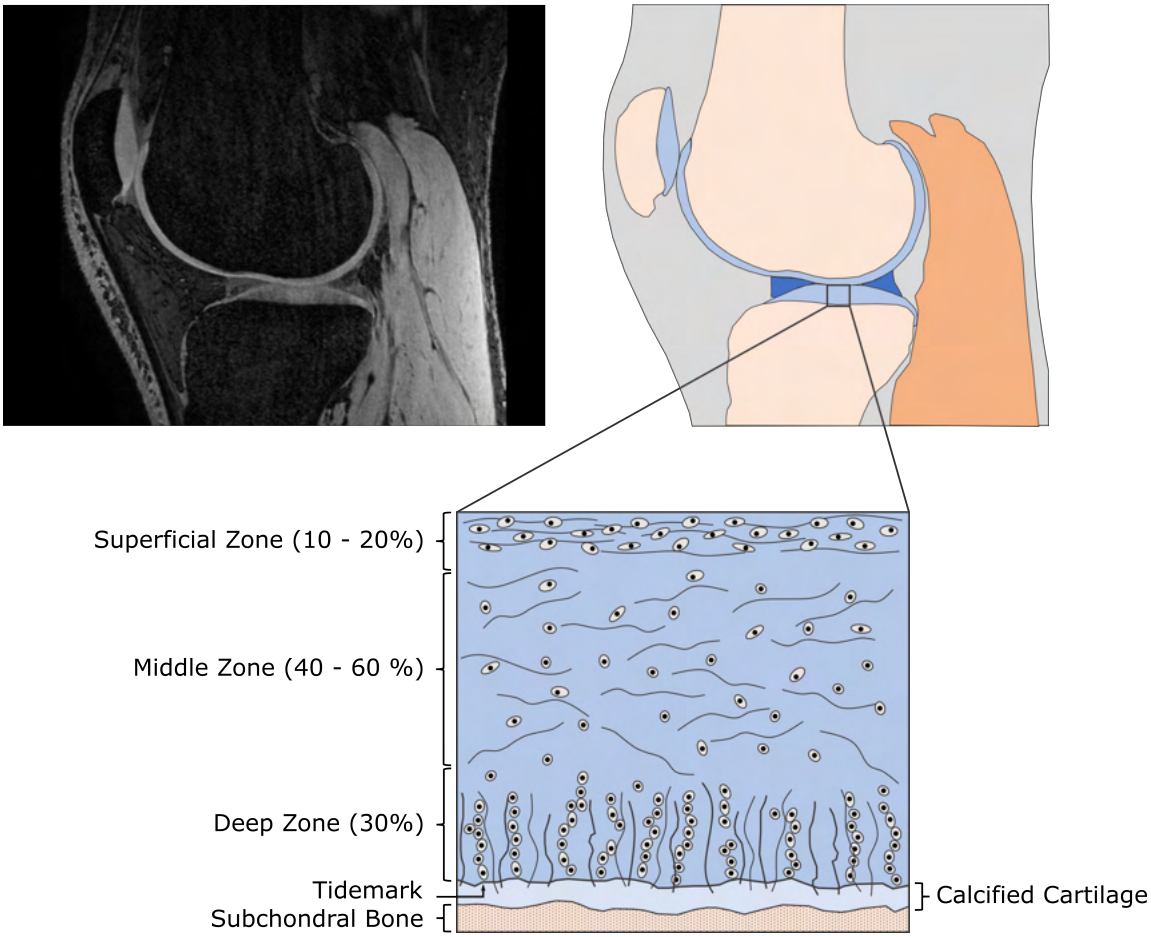

Figure 1.1: Illustration of the knee joint and structural components of articular cartilage.

An important function of the chondrocytes is the continuous synthesis and replacement of GAGs on the proteoglycans as these only have half-lives of weeks to years [9]. Collagen however, has a half-life of several decades and its production by chondrocytes slows down with aging [7, 9, 10]. Regular deformation of the cartilage structure and ultimately that of the ECM through moderate mechanical loading generates biochemical signals affecting the metabolic activity of the chondrocytes to maintain homeostasis throughout life [11–13].

1.1.2 Aging, Degeneration and Osteoarthritis

Osteoarthritis (OA) is a common degenerative disorder affecting one or multiple diarthrotic joints and a major cause of physical disability in the adult population [9, 14]. Symptomatically, OA is characterised by varying degrees of pain, stiffness, instability and functional impairment during normal daily activities and ultimately impairing quality of life [14–16]. OA is nowadays regarded as a disorder involving the entire joint. The pathophysiology of the disease and structural changes occurring during progression are cartilage loss and calcification, osteophyte formation, subchondral bone and meniscal modifications, and inflammation of the synovial joint lining [9, 14, 16, 17].

Early changes in osteoarthritic cartilage include the loss of macromolecular components such as proteoglycans and collagen. With the loss of negatively charged GAGs, the water content increases and the cartilage matrix swells up, resulting in the biomechanical abilities

of articular cartilage altering and load-bearing capabilities reducing [9, 11, 12, 16]. As OA progresses and cartilage matrix components deplete, the calcified cartilage and subchondral bone become exposed to forces during joint loading, which could lead to the development of bone marrow lesions [9, 16]. Both the dense subchondral cortical as well as the underlying, porous subchondral cancellous bone undergo compositional and structural alterations to adapt to OA-induced biomechanical variations Figure 1.2 [18]. Due to an increased bone remodelling rate in the early stages of OA, the subchondral cortical bone becomes thinner, and concurrently, cancellous bone is lost. In late-stage OA, the calcified cartilage increases and extends into the hyaline articular cartilage. Additionally, the cortical bone becomes thicker while the cancellous trabecular bone becomes sclerotic [19, 20].

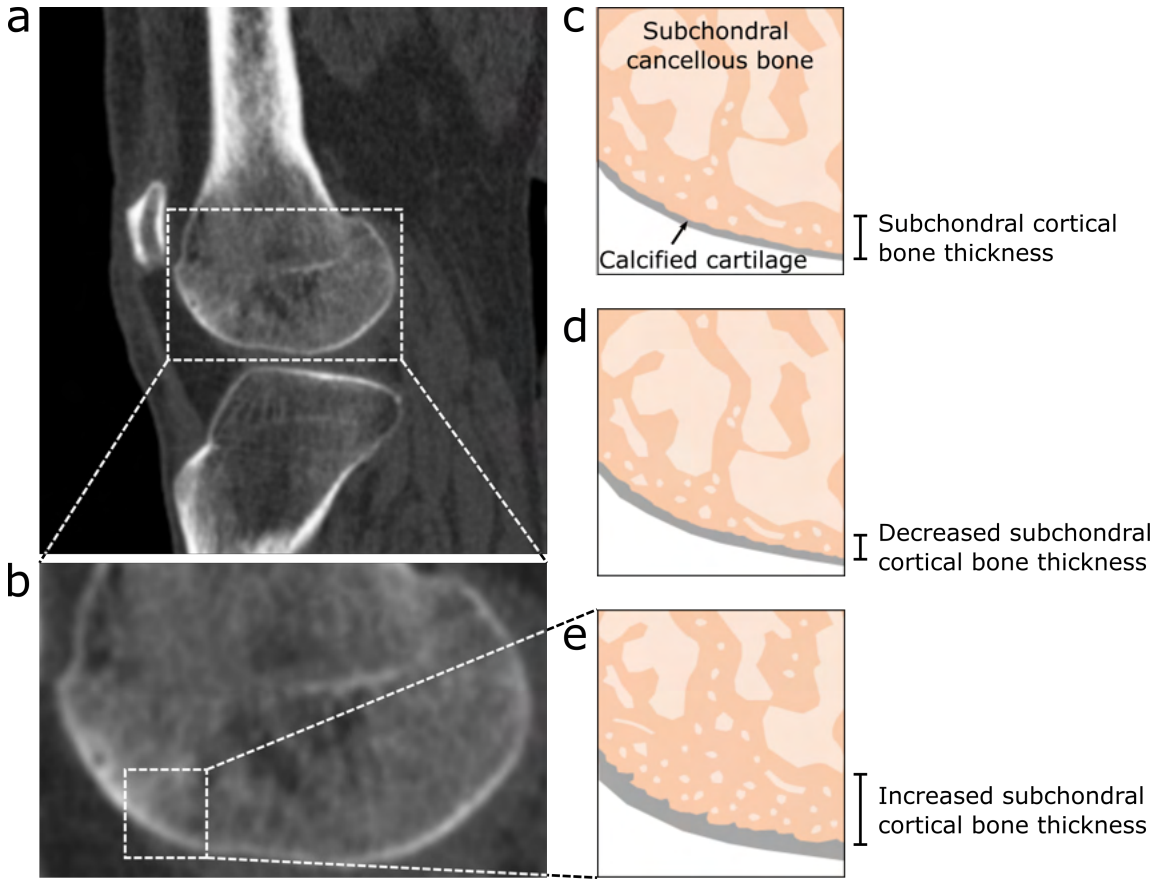

Figure 1.2: Model of subchondral bone remodelling during OA progression. (a) Sagittal CT image with (b) a close up of the femoral condyle. Illustration of (c) normal, (d) early OA and (e) late-stage OA subchondral bone composition and remodelling.

The prevailing risk factor for primary OA development is age, as cartilage metabolism and composition (water and macromolecular content/distribution) as well as chondrocyte activity change with advancing years [7, 21]. As a result, the load-bearing capabilities of the articular cartilage reduce and other structures, such as the subchondral bone, are consequently subjected to greater forces. Secondary OA results from another disorder or condition such as trauma or injury, obesity, genetics, knee malalignment and other abnormal joint biomechanics [9, 14, 22]. While long-term articular cartilage homeostasis is maintained while being subject to physiologic mechanical loads, repeated or continuous exposure of cartilage to abnormal loads (high-impact sports) or minimal loads (for example non-weight bearing after injury) can have degenerative effects on cartilage structure and ultrastructure [13, 22].

Currently, OA is treated using palliative approaches to treat symptoms such as pain evoked

by inflammation to improve mobility rather than target the disease itself [16, 23]. In most cases, treatment begins at advanced osteoarthritic stages where radiographic evidence of OA, such as joint space narrowing, is already present [9]. Treatments which can prevent OA, halt or slow down progression or even reverse the effects, are lacking at present [9, 23]. As OA progresses, a total joint replacement may be the final option to treat relentless severe pain. Developing new imaging methods and determining corresponding imaging biomarkers to identify the cartilage compositional state could help improve our understanding of the osteoarthritic process and aid the development of treatment strategies prior to irreversible joint damage [23].

1.1.3 Role of MRI in Imaging Articular Cartilage and Osteoarthritis

Clinical diagnosis of OA is based on patient-reported symptoms of joint pain and stiffness, and radiographic evidence of an osteophyte and joint space narrowing, indicating loss of cartilage and meniscal degeneration. Radiographic detection of OA means the disease is diagnosed at a later stage in the progression pathway and structural changes already exist.

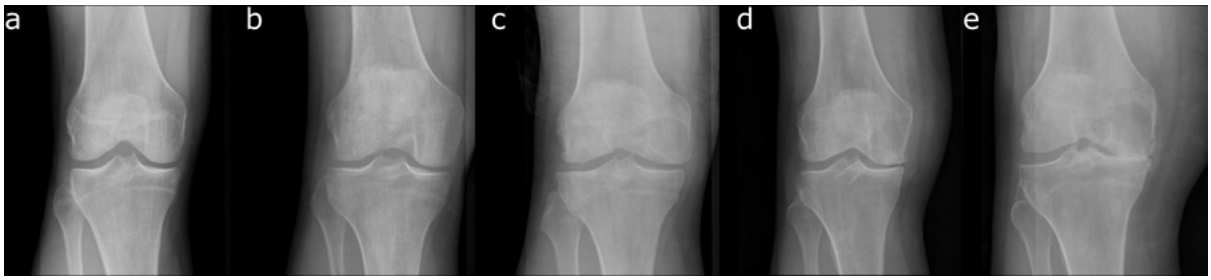

Figure 1.3: Radiographic progression of knee OA. Radiographs show a) normal appearance, b) doubtful to minimal joint space narrowing (JSN), c) mild JSN with formation of small osteophytes, d) moderate JSN and osteophyte development, and e) severe narrowing to complete loss of joint space with bone on bone contact.

The Kellgren and Lawrence (K-L) grading system [24] is most often used to classify radiographic knee OA. The system consists of five grades: None (0), Doubtful (1), Minimal (2), Moderate (3) and Severe (4). Structural knee joint changes used to grade the degree of OA are mainly the presence of osteophytes, the magnitude of joint space narrowing and an increased density of subchondral bone. Grade 0 indicates the certain absence, while grades $\geq$2 specify the definite appearance of radiographic changes associated with OA [24, 25].

Although the acquisition and classification of plain radiographs are cost effective and swiftly performed, they assume a linear progression of the disease and do not allow direct visualisation of all tissues involved in OA [26]. Recent research studies evaluating the longitudinal progression of knee OA, such as the Osteoarthritis Initiative, have incorporated magnetic resonance imaging (MRI) due to its ability to visualise bone and soft tissues and allow both the qualitative and quantitative assessment of articular cartilage and other joint tissues [27].

Qualitative MRI based on morphologic evaluation of OA-related abnormalities is most frequently used. The MRI Osteoarthritis Knee Score (MOAKS) system is one of the most commonly used semi-quantitative MR scoring system for knee OA [28]. MOAKS was developed from two existing scoring systems, the Whole Organ Magnetic Resonance Imaging Score (WORMS) and Boston Leeds Osteoarthritis Knee Score (BLOKS), in an effort to combine their advantages while also addressing their limitations [29, 30]. A common feature of all MR-based semi-quantitative scoring systems is that the knee is divided into multiple anatomical subre-

gions and graded according to the presence of various tissue abnormalities such as cartilage defects, meniscal tears, and bone marrow lesions [31]. Being able to evaluate various OA-related pathologies using conventional MR sequences that are widely and clinically available is a great advantage to the semi-quantitative assessment methods. Although the semi-quantitative scoring methods allow the evaluation of all tissues involved in OA, they are susceptible to inter-observer variability, dependent on the observers expertise and have been shown to be less sensitive to changes in cartilage and bone marrow lesions over a 2-year period compared to quantitative assessment [32].

Quantitative MRI methods can provide more detail on cartilage morphological and physiological state. Morphological quantifications typically include measurements of cartilage thickness, volume and surface area [33–35]. As the cartilage thickness only ranges over a few millimetres, these MR-based quantifications require images with high-resolution as well as high contrast differences between cartilage and other joint tissues to achieve accurate delineations. Consequently, gradient echo-based sequences with different methods of fat suppression are typically employed to measure cartilage morphological features. These include three-dimensional (3D) spoiled gradient echo (SPGR) sequences with fat-suppression or dual echo steady state (DESS) sequences with fluid-excitation.

In addition, several advanced quantitative MRI techniques have been developed to characterise the physiology (composition) of cartilage. These techniques include the measurement of transverse relaxation time (T_2) and the spin lattice relaxation time in the rotating frame ($T_{1\rho}$). Both T_2 and $T_{1\rho}$ are tissue-specific parameters that influence image contrast. Therefore, quantifying these relaxation times through T_2 and $T_{1\rho}$ mapping techniques can assist in tissue characterisation as well as determine differences between healthy and diseased tissues. Their potential lies particularly in the detection of early alterations in the biochemical composition of cartilage prior to any OA-induced morphological changes. T_2 relaxation time mapping is the most commonly used and well established technique for studying cartilage composition [36]. T_2 relaxation times have been shown to be sensitive to alterations in cartilage hydration as well as collagen integrity with prolonged T_2 relaxation times being associated with degenerated cartilage [37, 38]. $T_{1\rho}$ is increasingly being used assess cartilage composition and has been demonstrated to be sensitive to variations in the proteoglycan content of cartilage [39, 40]. $T_{1\rho}$ possesses a superior discriminatory ability to differentiate between various degrees of OA compared to T_2 [41, 42]. Although both T_2 and $T_{1\rho}$ relaxation time mapping techniques have been shown to be reliable and able to discriminate between normal and osteoarthritic cohorts, these techniques have not yet been adopted into routine clinical use [41]. At present, their use is mainly limited to academic institutions that have the required hardware and analysis expertise [43].

1.2 MRI Assessment of In Vivo Cartilage Response to Joint Loading Activities

Exercise programs have been suggested as a form of non-invasive treatment and management option of knee OA-related symptoms [44]. Increasingly more studies are evaluating the beneficial effects of different exercise programs for reducing knee pain and increasing mobility, however, these outcomes are usually based on objective measures [45]. By incorporating quantitative MRI measures of the effects of joint loading activities on cartilage structure and composition into these studies, a sensitive imaging biomarkers able to quantify cartilage health could

potentially be identified and further the development of appropriate treatment strategies for OA.

Numerous studies have examined the effects of daily living, different exercise types and in vivo mechanical loading on articular cartilage structure and composition to assess the repeatability and sensitivity of different quantitative MRI techniques. The purpose of this section is to provide a short summary of the current literature on the impact of various knee joint loading activities on articular cartilage structure (thickness or volume) and composition ($T_{1\rho}$ or T_2) assessed with in vivo MRI. A literature search strategy was formulated to identify all publications in Medline via PubMed and Scopus assessing diurnal-, exercise- and loading-related effects on knee articular cartilage structure and micro-structure using MRI. Studies involving participants with healthy, injured or diseased knee joints were considered. All identified publications were divided into three groups: Cartilage response to activities of daily living, exercise and mechanical joint loading.

Understanding the diurnal strains on articular cartilage is of great interest as these effects play an important part in maintaining healthy cartilage functionality and composition with cartilage degeneration potentially resulting from any alterations to normal, physiological cartilage loading. During normal daily loading, the cartilage structure and composition is varied through its compression resulting in water displacement from the cartilages extracellular matrix and altering its macromolecular content [46]. Morphological and compositional alterations in the whole cartilage due to diurnal effects were presented in most studies investigating these effects (Table 1.1) as non-significant. Nevertheless, it was highly recommended that the diurnal effects should be considered and avoided when conducting a longitudinal imaging study as significant focal deformations in the most weight-bearing cartilage regions were determined [46, 47]. Although the results from Li *et al* showed no significant difference in diurnal variations of $T_{1\rho}$ and T_2 relaxation values between morning and evening scans, the relaxation time measurements from the evening scans exhibited higher coefficients of variation than those determined during morning scans [48]. Consequently, they advised scanning participants in longitudinal studies in the morning to minimise diurnal-related variations of relaxation time measurements due to different daily activities of participants.

Studying the effects of different forms of static and dynamic joint loading exercise activities is important for determining the strain and magnitude of loading that promote healthy cartilage metabolism and when catabolic and cartilage degenerative effects begin to take over. The findings in the current studies (Table 1.2) suggest that short-term joint loading activities deform human articular cartilage independent of the cartilage state (healthy or diseased). While cartilage deformations seemed to be more focal and restricted to limited regions during more static activities such as a squat hold, the deformations seen during more dynamic exercises such as knee bends or running appeared to be more distributed over the entire cartilage surface [55, 58, 67, 71]. The structural and compositional response of cartilage suggests to depend more on loading type rather than the magnitude of load the whole joint is exposed to [69, 71]. The study results suggest good cartilage adaptation to static and dynamic loads and that cartilage possesses the ability to recover structurally from deformation within 15 - 90 min after joint loading, depending on loading type, cartilage region and cartilage state, but not necessarily on magnitude and duration of the joint loading activity [54, 60, 65]. With increasing age, the magnitude of cartilage deformation seems to become less, while sex, physical training status and duration of activity appears not to affect deformational behaviour [56, 58, 59]. A key confounder in imaging the effects of exercise on cartilage morphology and composition is, that the deformational behaviour of cartilage can only be determined shortly after and not immediately after exercise since time is needed to position the participant back on the MRI scanner before

Table 1.1: Summary of studies investigating the effects of daily living on articular cartilage using in vivo MRI.

Study	Cohort size	Cohort gender (M:F)	KL-Scores	Mean age (years)	Quantitative modality	Voxel size (mm^3)	Cartilage region	Loading type
Waterton et al., 2000 [47]	Normal: 12	Normal: 6:6	0	25.3 (range: 22-30)	Cartilage volume and thickness (SPGR)	0.55 x 0.55 x 1.56	Patellar	Diurnal (mainly standing activities)
Sitoci et al., 2012 [49]	Normal: 17	Normal: 9:7	0	M: 24.3 ± 3.2 F: 22.0 + 2.7	Cartilage volume (SPGR)	0.31 x 0.31 x 1.5	Patellar; Tibial	Nocturnal
Coleman et al., 2013 [50]	Normal: 10	Normal: 6:4	0	29.3 (range: 22-46)	Cartilage thickness (DESS)	0.31 x 0.31 x 1.0	Patellar; Tibial	Diurnal (daily activities)
Widmyer et al., 2013 [51]	Normal: 20	Normal: 16:4	0	Normal BMI: 30 ± 2 High BMI: 31 ± 2	Cartilage thickness (DESS)	0.29 x 0.29 x 1.0	Patellar; Tibiofemoral	Diurnal (daily activities)
Li et al., 2014 [48]	Normal: 6	Normal: 3:3	0	(range: 22-35)	Cartilage composition ($T_{1\rho}$, T_2)	0.55 x 1.10 x 4.0	Patellar; Tibiofemoral	Diurnal (daily activities)
Taylor et al., 2019 [52]	Normal: 7	Normal: 3:4	0	27 (range: 24-30)	Cartilage composition ($T_{1\rho}$)	0.55 x 0.55 x 3.0	Patellar; Tibiofemoral	Diurnal (daily activities)

Table 1.2: Summary of studies investigating the effects of different exercise regimes on articular cartilage using in vivo MRI.

Study	Cohort size	Cohort gender (M:F)	KL-Scores	Mean age (years)	Quantitative modality	Voxel size (mm^3)	Cartilage region	Loading magnitude
Eckstein et al., 1998 [53]	Normal: 8	Normal: 5:3	0	(range: 23-32)	Cartilage volume (SPGR)	0.29 x 0.29 x 2.0	Patellar	50 knee bends
Eckstein et al., 1999 [54]	Normal: 7	Normal: 4:3	0	(range: 23-32)	Cartilage volume (SPGR)	0.29 x 0.29 x 2.0	Patellar	Repetitive sets of 50 knee bends; 100 knee bends
Eckstein et al., 2000 [55]	Normal: 12	Normal: 6:6	0	25.3 (age range: 22-30)	Cartilage volume and thickness (SPGR)	0.31 x 0.31 x 1.5	Patellar	30 deep knee bends; 90° squatting
Hudelmaier et al., 2001 [56]	Normal: 95 (young); 30 (elderly)	Normal: 49:46 (young); 15:15 (elderly)	0	Young: (range: 20-30); Elderly: (range: 50-78)	Cartilage thickness (SPGR)	0.31 x 0.31 x 1.5	Patellar	30 deep knee bends
Liess et al., 2002 [57]	Normal: 20	Normal: 16:4	0	28 ± 6	Cartilage thickness (FS MSME); Cartilage composition (T2)	0.31 x 0.31 x 1.5	Patellar	60 knee bends
Eckstein et al., 2005 [58]	Normal: 12 (patellar); 10 (tibiofemoral); 7 (weight-lifters); 7 (sprinters); 14 (untrained)	Normal: 6:6 (patellar); 5:5 (tibiofemoral); N/A (weight-lifters); N/A (sprinters); 14:0 (untrained)	0	Patellar: (range: 23-30); Tibiofemoral: (range: 18-37); Weight-lifters: 23.0 ± 3.2; Sprinters: 29.3 ± 4.2; Untrained: 24.9 ± 1.8;	Cartilage volume (SPGR)	0.31 x 0.31 x 1.5	Patellar; Tibiofemoral	Squatting; Walking; Running (200 m); Stepping; Cycling; 30 deep knee bends; Drop landings (10 vertical drops from 40 cm height);
Kessler et al., 2006 [59]	Normal: 48	Normal: 48:0	0	38 ± 14	Cartilage volume (SPGR)	0.31 x 0.31 x 1.5	Patellar; Tibial	Running (5, 10 and 20 km)
Kessler et al., 2008 [60]	Normal: 20	Normal: 20:0	0	38 ± 14	Cartilage volume (SPGR)	0.31 x 0.31 x 1.5	Patellar; Tibial	Running (20 km)
Mosher et al., 2010 [61]	Normal: 22 (athletes) & 15 (controls)	Normal:	0		Cartilage thickness (SPGR); Cartilage composition (T2)	0.33 x 0.33 x 4.0	Tibiofemoral	Running (30 min)
Farrokhi et al., 2011 [62]	Normal: 10 PFP: 10	Normal: 0:10 PFP: 0:10	N/A	Normal: 27.0 ± 4.4 PFP: 27.7 ± 4.3	Cartilage thickness (SPGR); Cartilage composition (T2)	3D-SPGR: 0.31 x 0.31 x 2.0 T2: 0.42 x 0.83 x 4.0	Patellar	50 deep knee bends
Niehoff et al., 2011 [63]	Normal: 14	Normal: 7:7	0	23.1 ± 2.1 (range: 18-26)	Cartilage volume and thickness (SPGR)	0.31 x 0.31 x 1.5	Patellar; Tibiofemoral	Drop landings (100 vertical drops from 73 cm height); Running (4 km)
Subburaj et al., 2012 [64]	Normal: 20	Normal: 10:10	0	28.8 (range: 22-35)	Cartilage thickness (SPGR); Cartilage composition (T1rho)	3D-SPGR: 0.27 x 0.27 x 1.0 T1rho: 0.55 x 0.55 x 3.0	Patellar; Tibiofemoral	Running (30 min)
Van Ginckel, 2013 [65]	Normal: 18 OA: 18	Normal: N/A OA: 12:6	Normal: 0 OA: 1-2	Normal: 43.0 OA: 54.5	Cartilage volume (DESS)	0.36 x 0.36 x 0.70	Tibiofemoral	30 Squats

Sutter et al., 2015 [66]	Normal: 8	Normal: 8:0	0	26.3 (range: 24-30)	Cartilage thickness (DESS)	0.31 x 0.31 x 1.0	Tibiofemoral	60 single-legged hops
Horng et al., 2015 [67]	Normal: 10	Normal: 5:5	0	23 ± 1	Cartilage thickness (3D FLASH)	0.31 x 0.31 x 1.5	Patellar; Tibiofemoral	Kneeling (10 min); Squatting (10 min); Heel-Sitting (10 min); 50 knee bends
Lad et al., 2016 [68]	Normal: 8	Normal: 4:4	0	25.4 (range: 22-30)	Cartilage thickness (DESS)	0.31 x 0.31 x 1.0	Tibiofemoral	Walking (20 min)
Gatti et al., 2017 [69]	Normal: 15	Normal: 15:0	0	(range: 18-35)	Cartilage composition (T2)	0.63 x 0.63 x 3.0	Tibiofemoral	Running; Bicycling
Liu et al., 2017 [70]	Normal: 8	Normal: 4:4	0	25.4 (range: 22-30)	Cartilage thickness (DESS)	0.31 x 0.31 x 1.0	Tibial	Walking (20 min)
Chen et al., 2017 [71]	Normal: 23	Normal: 11:12	0	25 (range: 23-30)	Cartilage composition (T1rho, T2)	T1rho: 0.49 x 0.73 x 3.0 T2: 0.55 x 0.88 x 2.5	Patellar; Tibiofemoral	Walking; Running: Stair activity
Owusu-Akyaw et al., 2018 [72]	Uni-ACL: 8	Uni-ACL: 8:0	N/A	31 (range: 21-47)	Cartilage thickness (DESS)	0.31 x 0.31 x 1.0	Patellofemoral	60 single-legged hops
Ho et al., 2018 [73]	Normal: 10 OA: 9	Normal: 5:5 OA: 5:4	Normal: ≤1 OA: ≥2	Normal: 55.0 ± 1.8 OA: 55.6 ± 4.5	Cartilage thickness (SPGR)	0.35 x 0.35 x 2.0	Femoral	Walking (30 min)
Collins et al., 2018 [74]	Normal: 15	Normal: 8:7	0	Normal BMI: 30 (range: 23-43 High BMI: 32 (range: 22-45)	Cartilage thickness (DESS); Cartilage composition (T1rho)	DESS: 0.31 x 0.31 x 1.0 T1rho: 1.10 x 0.50 /x 3.0	Tibiofemoral	Walking (20 min)
Sutter et al., 2019 [75]	Uni-ACL: 8	Uni-ACL: 8:0	N/A	31 (range: 21-47)	Cartilage thickness (DESS)	0.31 x 0.31 x 1.0	Tibiofemoral	60 single-legged hops

Table 1.3: Summary of studies investigating the effects of mechanical joint loading on articular cartilage using in vivo MRI.

Study	Cohort size	Cohort gender (M:F)	KL-Scores	Mean age (years)	Quantitative modality	Voxel size (mm^3)	Cartilage region	Loading magnitude
Nag et al., 2004 [76]	Normal: 26	Normal: 14:12	0	52.0 ± 15.0	Cartilage composition (T2)	0.59 x 0.97 x 3.0	Tibiofemoral	60 kg
Nishii et al., 2008 [77]	Normal: 22	Normal: 9:13	0	25 (range: 18-43)	Cartilage composition (T2)	0.23 x 0.23 x 3.0	Tibiofemoral	50 % body weight
Souza et al., 2010 [78]	Normal: 10 OA: 20	Normal: 0:10 OA: 0:20	Normal: 0 OA: 2-3	Normal: 52.9 ± 6.5 OA: 57.1 ± 4.7	Cartilage composition (T1rho, T2)	0.50 x 0.50 x 3.0	Tibiofemoral	50 % body weight
Shin et al., 2011 [79]	Normal: 11 OA: 10	Normal: 0:11 OA: 0:10	Normal: 0 OA: 3	Normal: 51.8 ± 6.7 OA: 57.4 ± 4.2	Cartilage-to-cartilage contact area (SPGR, T2w-FSE)	3D-SPGR: 0.27 x 0.27 x 1.5 T2w-FSE: 0.36 x 0.73 x 2.0	Tibiofemoral	50 % body weight
Cotofana et al., 2011 [80]	Normal: 11 OA (KL2): 10 OA (KL3): 9	Normal: 0:11 OA (KL2): 0:10 OA (KL3): 0:9	Normal: 0 OA: 2-3	Normal: 52.0 ± 6.8 OA (KL2): 56.0 ± 5.4 OA (KL3): 57.9 ± 4.1	Cartilage thickness (SPGR)	0.31 x 0.31 x 1.5	Tibiofemoral	50 % body weight
Subburaj et al., 2012 [81]	Normal: 10 OA: 20	Normal: 0:10 OA: 0:20	Normal: 0 OA: 2-3	(range: 42 – 68)	Cartilage-to-cartilage contact area (3D-SPGR) Cartilage composition (T1rho, T2)	0.31 x 0.31 x 1.5	Tibiofemoral	50 % body weight
Souza et al., 2014 [82]	Normal: 93 OA: 44	Normal: 39:54 OA: 17:27	Normal: 0-1 OA: 2-4	Normal: 49.5 ± 1.9 OA: 57.4 ± 3.0	Cartilage composition (T1rho, T2)	0.55 x 1.10 x 4.0	Tibiofemoral	50 % body weight
Schoenbauer et al., 2015 [83]	Normal: 9 MACT: 10	Normal: 6:3 MACT: 6:4	Normal: N/A MACT: N/A	Normal 31 ± 7 MACT: 40 ± 8	Cartilage composition (T2)	0.25 x 0.25 x 3.0	Tibiofemoral	120 N
Lange et al., 2017 [84]	Normal: 10	Normal: 10:0	0	Normal: 27 ± 2	Cartilage composition (T1rho, T2)	0.60 x 0.60 x 3.0	Patellofemoral	20 kg, 40 kg
Lange et al., 2019 [85]	Normal: 15	Normal: 15:0	0	Normal: 28.5 ± 2.2	Cartilage thickness and cartilage-to-cartilage contact area (SPGR)	0.40 x 0.40 x 0.5	Patellofemoral	200 N, 400 N

being able to acquire the data.

An MR compatible loading device can assist in determining the immediate effects that mechanical loading of the knee joint has on cartilage morphology and composition. The majority of the current studies using loading devices (Table 1.3) applied a load of about 50% of the participants body weight as this is thought to emulate the magnitude of static loading of the knee joint present during normal standing. Additionally, joint loading was typically applied for 8-20 minutes prior to post-loading MR imaging. The findings from Nag and colleagues [76] suggest that healthy ageing of cartilage does not influence the response ability to mechanical loading. However, degenerated cartilage of OA patients showed slightly less weight-bearing capabilities through greater deformational and compositional changes compared to healthy cartilage during loading [80, 82]. The medial cartilage compartments seem to be more involved in the weight-bearing and -distribution process during loading showing larger compositional variations and an increase in cartilage-to-cartilage contact area [78, 79, 81]. Additionally, results from Schoenbauer *et al* suggest plateau-like cartilage adaptation to static mechanical loads with increasing load duration [83].

To summarise, high-resolution 3D gradient echo sequences have shown to provide reliable measures of cartilage volume and thickness, allowing the assessment of cartilage compression and deformation in response to different types and magnitudes of joint loading. Additionally, quantitative T_2 and $T_{1\rho}$ relaxation time mapping techniques suggest being reproducible and permit the assessment of cartilage compositional alterations following joint loading.

1.3 Deep Convolutional Neural Networks for Knee Joint Tissue Segmentation

Most quantitative MR imaging methods used for assessing musculoskeletal joint tissues have not been able to make the step into clinical practice due to the large burden placed upon the observes to manually validate these methods. Developing tools that can automatically analyse the health state of joint tissues fast, reliable and accurate are therefore desirable. Additionally, such tools could aid in evaluating large scale longitudinal knee imaging datasets such as the Osteoarthritis Initiative (OAI) or the Applied Private-Public partneRship enabling OsteoArthritis Clinical Headway (APPROACH) studies.

The motivation behind deep learning comes from the inability of simple machine learning algorithms extracting hand-crafted features to generalise large, high-dimensional datasets [86]. Convolutional neural networks, or CNNs, are deep neural networks designed to extract important texture and intensity features from data grids such as images or time-series. CNNs have shown tremendous potential in overcoming the time-consuming, demanding and variability-prone process of manual tissue segmentation from medical images [87]. The CNN architectures consist of different types of layers that either perform a fixed mathematical function or a transformation on the inputs. The building blocks of almost all CNNs are:

- Convolutional layers are the core parts and the so-called feature extractors of any CNN. Every input passing through a convolutional layer will have a defined number of small n x n filters slide or convolve over its grid and output the dot product of each grid and filter entry. These layers take advantage of sparse connectivity within the input image by constraining the output entries on a spatially local subarea of the input. The output of a convolutional layer is referred to as a feature map. The number of feature map

representations of the input is determined by the number of filters convolving over the input. The size of the output is defined by the stride with which the filter is slid across the input as well as the size of zero-padding around the inputs boarder.

- Pooling layers, which help reduce the dimensionality and the computational time of the network by outputting a summary statistic of a spatially local subset of the input. The most common statistical pooling operations are max pooling (only forward the maximum value of the input subset) and average pooling (forward the average value of the input subset). This layer helps preserve the important feature information of the image, while less important features are removed.

- Activation layers, which are typically applied after every convolutional layer to introduce non-linearity into the network and allow it to solve more complex problems. The most frequently used activation functions are the $sigmoid$ function, the $tanh$ function as well as the $ReLU$ (Rectified Linear Unit) and $leakyReLU$ functions [88].

In supervised learning, the training phase of the CNN consists of the minimisation of a pre-defined objective function, also known as the loss function $\mathcal{L}$, by using a pre-annotated dataset and comparing the ground truth segmentations with the networks predicted segmentations. In segmentations tasks, this loss function is usually the Cross Entropy or Binary Cross Entropy loss, or loss functions based on segmentation evaluation metrics such as the Dice similarity coefficient [89].

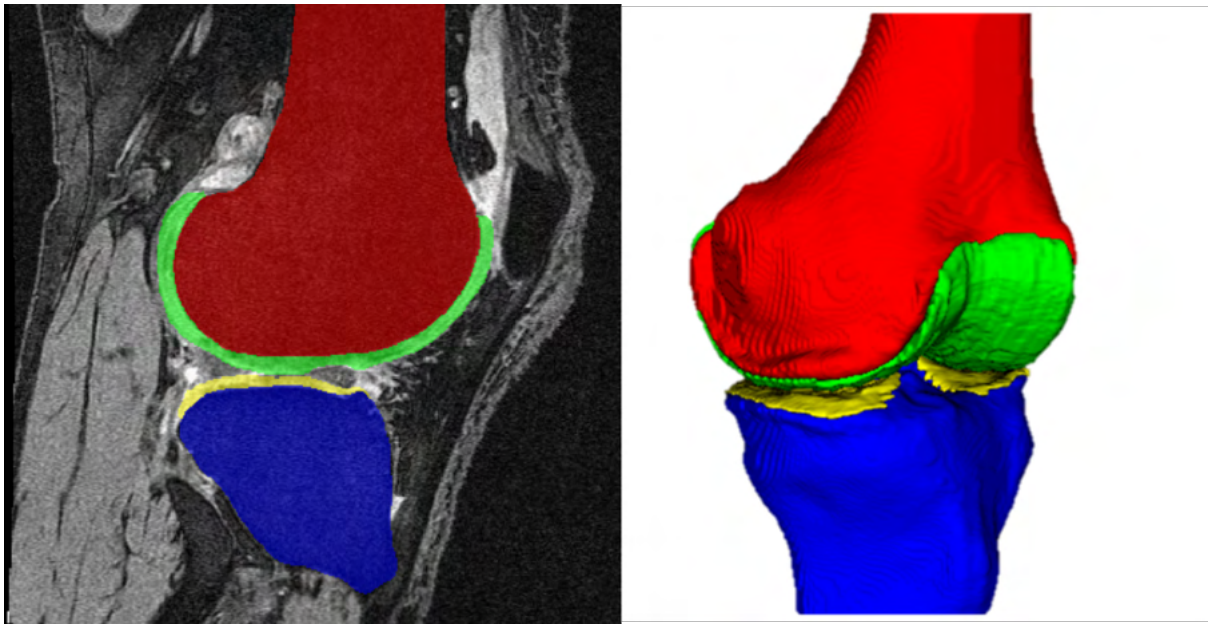

Figure 1.4: Sagittal DESS image overlayed with manual cartilage and bone segmentations (left) and the corresponding volumetric representation of the bone and cartilage surfaces (right).

Two layers used to typically speed up and stabilize training of deep neural networks are the Batch Normalisation and Dropout layers. The Batch Normalisation layer enables stability during learning by decreasing the influence of outlying large weights that can impact the training process [90]. These layers are typically placed after convolutional layers to produce normalised activation maps by subtracting the mean and dividing by the standard deviation for each training batch. By normalizing, one can use larger training rates during loss optimisation

which speeds training. This also makes the network less dependent on careful parameter initialisation. The Dropout layer changed the concept of learning all input entries to learning only from a random fraction of the input entries in the network in each training iteration [91]. Therefore, during loss optimisation, a different, slightly smaller network is trained in each iteration. Dropout layers consequently act as regularisers as they force the network to learn more robust features.

To assess the accuracy of the network-generated segmentation map compared to the ground truth, different evaluation metrics are used. Overlap-based metrics such as the Dice similarity coefficient [89] or Jaccard similarity coefficient [92] evaluate the accuracy of spatial overlap of generated and true segmentation. Volume-based metrics such as the volume overlap error assess volumetric overlap between ground truth and automated segmentations. Lastly, distance-based metrics such as the average symmetric surface distance evaluate the similarities between automated and manual segmentation contours. They measure the distance between each voxel from the manual segmentation contours to the nearest voxel on the contours of the automated segmentation and vice versa.

The most widely used CNN for automated image segmentation is the U-Net introduced by [93]. This CNN consists of a contracting path and an expansive path, which gives it the u-shaped architecture. In the contracting (encoding) path, an input image is increasingly down-sampled for coarse feature detection. In the ensuing expansive (decoding) path, the data is up-sampled to allow detailed feature localisation. Together with additional skip-connections, in which features from the contracting path are concatenated to features from the expanding path, high-resolution outputs (segmentations) are attained.

Image-to-image translation CNNs such as Generative Adversarial Networks (GANs) as well as their conditional variant (cGAN) have also recently been exploited for segmentation of musculoskeletal MRIs [94, 95]. GANs involve two CNNs, a generative and a classification network that are trained competitively and simultaneously towards optimising a loss function [96]. The generative network, typically a U-Net, focusses on generating realistic translations of the ground truth training images, while the classification / discriminator network decides if both the generated and ground truth images are accurate representations of each other, i.e. come from the same data distribution. While GANs typically generate images from random noise inputs, cGANs additionally receive an image as input which to translate to the desired output [97].

Further details on the architectures and trainings of U-Nets and cGANs for segmentation of knee joint tissues from MRIs will be provided in Chapters 4 - 6 of this thesis and therefore not be discussed here.

Chapter 2

Quantitative Magnetic Resonance Imaging

2.1 Principles of Nuclear Magnetic Resonance

Magnetic resonance imaging (MRI) is a non-ionising, cross-sectional imaging method in medicine with which metabolic and functional information of the human body can be obtained. Atomic nuclei that are made up of an odd number of nucleons possess an intrinsic angular momentum or spin **J** and a therewith connected magnetic dipole moment:

$$\boldsymbol{\mu} = \gamma \cdot \mathbf{J}. \tag{2.1.1}$$

Here, γ is the gyromagnetic ratio, a characteristic constant of each type of nucleus. The magnitude of the angular momentum is given by:

$$|\mathbf{J}| = \hbar \left[I(I+1) \right]^{\frac{1}{2}} \tag{2.1.2}$$

where $\hbar$ is Plank's constant h divided by 2π and I is the nuclear spin quantum number of the nucleus under investigation. The values of γ and I for nuclei frequently used in MRI are listed in Table 2.1.

Table 2.1: Gyromagnetic ratio and spin quantum number of nuclei used in MRI.

Isotope	^{1}H	^{13}C	^{19}F	^{23}Na	^{31}P
Gyromagnetic ratio, γ (MHz/T)	42.58	10.71	40.08	11.26	17.24
Spin quantum number, I	1/2	1/2	1/2	3/2	1/2

Clinical MRI mainly includes the imaging of hydrogen atoms (or protons, ^{1}H) since the human body consists of 70% water. Hydrogen has a half-integer nuclear spin and the largest gyromagnetic ratio. In a homogeneous magnetic field $B_0 = B_z$, a Zeeman splitting of the spins into two spin states is observed. The associated energy of the spins is given by

$$E = -\mu_z \cdot B_0 \tag{2.1.3}$$

where $\mu_z = \gamma J_z = \gamma \hbar m$ and $m = (-\frac{1}{2}, +\frac{1}{2})$. Therefore, two equidistant energy states exist for protons:

$$E = -\gamma \hbar m B_0 = \pm \frac{1}{2} \gamma \hbar B_0 \tag{2.1.4}$$

with the energy difference ΔE between the states being $\Delta E = \gamma \hbar B_0$. Within an applied, static magnetic field, the protons do not perfectly align with the magnetic field lines, but rotate around them at an angle α in a gyro-shape motion. The frequency at which this precession takes place depends on the magnetic field strength B_0 and is given by:

$$\omega_L = \frac{\Delta E}{\hbar} = \gamma \cdot B_0. \tag{2.1.5}$$

This frequency is known as the Larmor frequency.

In addition, the spins tend to align either parallel (energetically more favourable, lower energy state $E_\uparrow$) or anti-parallel (higher energy state $E_\downarrow$) with the magnetic field lines. The population probability of the spins between these two energy states is determined by the Boltzmann equation:

$$\frac{N_\uparrow}{N_\downarrow} = \exp\left(-\frac{E_\uparrow - E_\downarrow}{k_B T}\right) = \exp\left(-\frac{\Delta E}{k_B T}\right) = \exp\left(-\frac{\gamma \hbar B_0}{k_B T}\right) \tag{2.1.6}$$

where k_B is the Boltzmann constant and T is the absolute temperature in Kelvin.

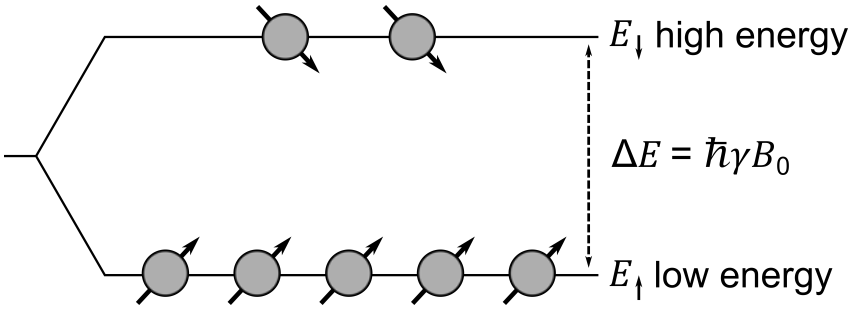

Figure 2.1: Boltzmann distribution of the proton spins within an applied static magnetic field. The spins are either oriented parallel (spin up, lower energy state $E_\uparrow$) or anti-parallel (spin down, higher energy state $E_\downarrow$) with the magnetic field.

At thermal equilibrium, the lower energy state $E_\uparrow$ is occupied slightly greater than the higher energy state $E_\downarrow$ ($N_\uparrow > N_\downarrow$). Given the large number of protons in the human body, this slight preference leads to a measurable magnetisation M_0 in the z-direction, i.e. parallel to the applied magnetic field.

2.2 Signal Generation

By emitting an alternating magnetic field (radio-frequency RF pulse) with a frequency that corresponds exactly to the Larmor frequency, energy is supplied to the system. This leads to

a tilting or tipping of the nuclear spins and the longitudinal magnetisation from the z-direction into the transverse, x-y plane. The flip angle α by which the longitudinal magnetisation is tipped from the z-axis depends on the amplitude B_1 and the duration t_p of the RF pulse.

$$\alpha = \gamma B_1 t_p \tag{2.2.1}$$

Emitting a $\frac{\pi}{2}$ RF pulse causes the magnetisation to be tipped by 90° into the x-y plane. Here the magnetisation precesses around the z-axis (transverse magnetisation M_{xy}) and induces a measurable voltage within a receiving coil, which measures the MR signal. The amplitude of this signal quickly decays exponentially to zero, since the protons quickly dephase due to energy exchange with nearby nuclear system caused by magnetic field inhomogeneities. This signal is known as Free Induction Decay (FID). Emitting an RF pulse either twice the strength or the duration of the $\frac{\pi}{2}$ pulse inverts M_0 by 180° and is called a π or inversion pulse. These RF pulses can be used as an excitation pulse (tipping M_0 away from B_0) which causes the initial signal or refocusing pulse (bringing the magnetisation back towards B_0) to refocus signal that has previously been excited.

2.3 Relaxation

2.3.1 T_1 and T_2 Relaxation

Two relaxation processes convert an induced transverse magnetisation back into the intial and stable longitudinal magnetisation at thermal equilibrium. These two processes are called longitudinal (T_1) relaxation and transverse (T_2) relaxation.

The longitudinal T_1 relaxation causes the recovery of the longitudinal magnetisation M_z through the interaction of the spin system (in the form of energy transfer) with the surrounding macromolecules (lattice). Therefore, it is also referred to as the spin-lattice relaxation. T_1 relaxation occurs due to the tumbling of the surrounding molecules at the Larmor frequency, generating fluctuating magnetic fields and stimulating a loss of energy as the the spins return to the lower energy state in thermal equilibrium. The time required to recover 63% of the original longitudinal magnetisation M_0 is defined as the T_1 relaxation time.

The transverse T_2 relaxation causes a decrease in the transverse magnetisation M_{xy} due to spin-spin interaction resulting in the loss of the phase coherence of the precessing spins in the transverse plane. Therefore, it is also referred to as the spin-spin relaxation. The T_2 relaxation time refers to the time for the transverse magnetisation to decay to 37% of M_0. While T_1 and T_2 relaxation take place simultaneously, M_{xy} dephasing is significantly faster than M_z recovery. Additional local magnetic field inhomogeneities (ΔB_0) caused by different macromolecular environments result in an additional signal decay of the transverse magnetisation. Therefore, the overall rate of dephasing is increased and described by the T_2^* relaxation time. T_2^* is defined as:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B_0 \tag{2.3.1}$$

With the help of the Bloch equations, macroscopic alterations in the magnetisation during excitation and relaxation can be described. If **M** describes the summation of all magnetic

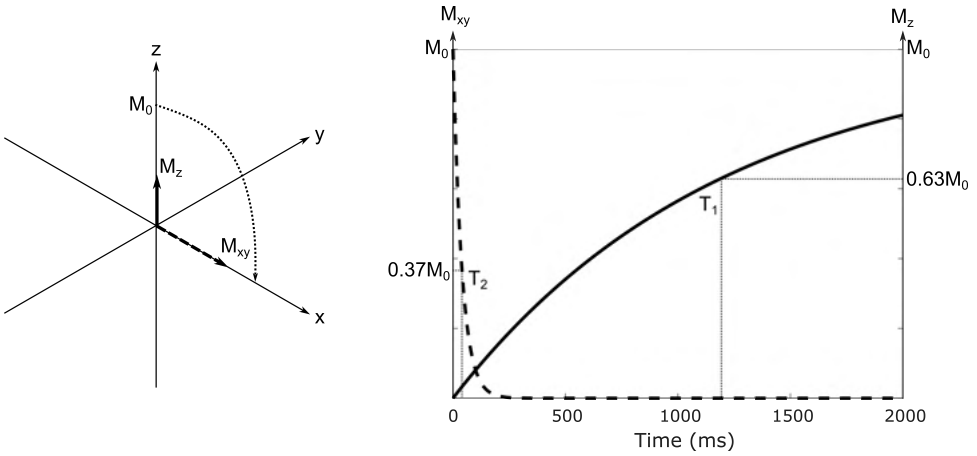

Figure 2.2: Longitudinal and and transverse relaxation simulated for typical relaxation time values for articular cartilage (T_1 = 1200 ms and T_2 = 40 ms).

moments of the spin system, then the change in **M** over time due to the experienced torque of the spins in an external magnetic field **B** can be defined as:

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B} \tag{2.3.2}$$

where **M** is made up of the transverse magnetisation components, M_x and M_y, and the longitudinal component, M_z. Similarly, **B** encompasses the static magnetic field along the z-axis,

$$B_z = B_0 \tag{2.3.3}$$

and B_x and B_y which are the rotating B_1 magnetic fields along the x and y direction:

$$B_x = B_1 \cos(\omega t), \tag{2.3.4}$$

$$B_y = -B_1 \sin(\omega t). \tag{2.3.5}$$

By expanding the vector product of Equation 2.3.2 and including the T_1 and T_2 relaxation processes as first-order processes, the directional components of **M** can be described as:

$$\frac{dM_x}{dt} = \gamma \left(M_y B_0 + M_z B_1 \sin(\omega t) \right) - \frac{M_x}{T_2}, \tag{2.3.6}$$

$$\frac{dM_y}{dt} = \gamma \left(-M_x B_0 + M_z B_1 \cos(\omega t) \right) - \frac{M_y}{T_2}, \tag{2.3.7}$$

$$\frac{dM_z}{dt} = \gamma \left(-M_x B_1 \sin(\omega t) + M_y B_1 \cos(\omega t) \right) - \frac{M_z - M_0}{T_1}. \tag{2.3.8}$$

If a 90° RF excitation pulse is applied, where immediately after the pulse, $M_{x,y} = M_0$ and $M_z = 0$, the solutions to these differential equations are:

$$M_x(t) = M_0 \sin(\omega t) e^{-\frac{t}{T_2}} \tag{2.3.9}$$

$$M_y(t) = M_0 \cos(\omega t) e^{-\frac{t}{T_2}} \tag{2.3.10}$$

$$M_z(t) = M_0 \left[1 - e^{-\frac{t}{T_1}}\right]. \tag{2.3.11}$$

While the transverse magnetisation components M_x and M_y precess at the Larmor frequency and decay back to equilibrium ($M_{x,y}$ = 0) at a rate 1/T_2, the longitudinal magnetisation M_z exponentially recovers back to its equilibrium (M_z = M_0) at a rate 1/T_1.

2.3.2 Relaxation in the Rotating Frame

$T_{1\rho}$ is the relaxation time of transverse magnetisation under the influence of an external spin-lock RF pulse. $T_{1\rho}$ is also regarded as the time constant for spin-lattice relaxation in the rotating frame.

After tipping the magnetisation into the transverse plane using a 90° RF excitation pulse, a continuous wave spin-lock RF pulse B_{SL} is applied, locking the spins in the transverse plane. While the spin-lock pulse is applied, the spins precess around B_{SL} rather than B_0 with an angular frequency

$$\omega_{SL} = \gamma \cdot B_{SL}. \tag{2.3.12}$$

and the magnetisation decays with the time constant $T_{1\rho}$ in a new composite rotating frame of the spin lock pulse:

$$M(TSL) = M_0 e^{-\frac{TSL}{T_{1\rho}}} \tag{2.3.13}$$

where TSL is the spin-lock time.

Consequently, in contrast to the conventional spin-lattice relaxation (T_1) that is sensitive to fast motion interactions near the Larmor frequency (MHz range), $T_{1\rho}$ can probe low frequency biological interactions (few hundred Hz to kHz) in macromolecular environments. Therefore, by varying the amplitude of the spin-lock pulse B_{SL}, the macromolecular interactions at different frequencies can be probed with $T_{1\rho}$ imaging.

2.4 The MR System

The MR system typically comprises of three main components: (1) a large static magnetic field, (2) a radio-frequency (RF) coil system to transmit energy and receive the MR signal, and (3) gradient coils to generate spatially varying magnetic fields and encode the MR signal.

2.4.1 Magnet

The static magnetic field in clinical systems is typically produced by a superconducting coil system made of an niobium-titanium alloy and immersed in liquid helium at about 4 K temperature. The low temperatures are fundamental to achieve superconductivity and simultaneously zero electric resistance in the coil windings avoiding heating. Clinical MR systems have magnetic field strength of typically 1.5 T or 3.0 T. Strong magnetic fields are required to obtain the necessary nuclear polarisation introduced in Section 2.1 to allow MR imaging.

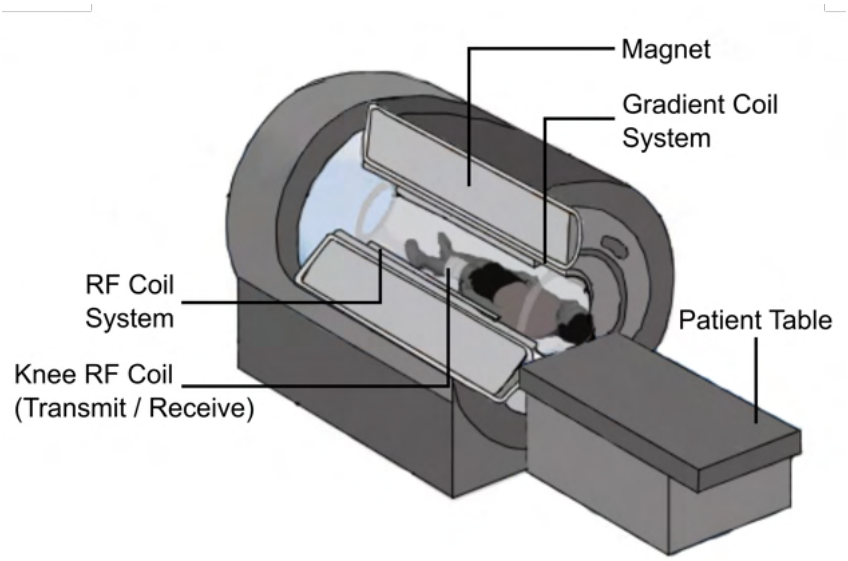

Figure 2.3: Illustration of an MR system. Example of patient undergoing knee MRI with a dedicated transmit / receive knee RF coil.

An important requirement to the static magnetic field is that it should be highly uniform to allow accurate spatial encoding. Field inhomogeneities are minimised by active shim coils while additional superconductive shielding coils reduce the effect of external fields on the environment.

2.4.2 Gradient Coil System

For the exact spatial localisation of the MR signal in each voxel, three linear, spatially-varying magnetic field gradients in the x, y and z directions are superimposed onto the main static magnetic field. These magnetic field gradients are produced by three pairs of orthogonal gradient coils. For the x and y direction, the coil pairs are oriented in the Golay configuration and for the z direction, a Maxwell coil pair is used.

Slice Selection

For the slice selection, a gradient in the z-direction is superimposed on the main magnetic field during the application of an RF excitation pulse, which is intended to induce a flipping of the nuclear spins and thereby a build-up of the transverse magnetisation. This leads to different Larmor frequencies of the nuclear spins in each z-plane

$$\omega(z) = \gamma \left(B_0 + z G_{SS} \right) . \tag{2.4.1}$$

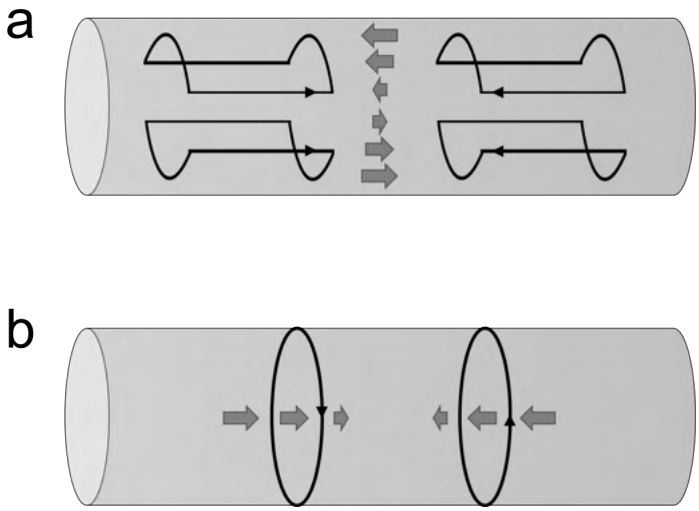

Figure 2.4: Different gradient coil configurations are used to generate magnetic field gradients in x, y (a) Golay configuration) and z direction (b) Maxwell coil pair).

In order to be able to map a certain slice, the centre frequency of the RF excitation pulse must correspond exactly to the specific Larmor frequency of the spins in the selected slice. The thickness of the slice Δz is defined by the amplitude of the slice select gradient G and the bandwidth of the sinc-shaped RF excitation pulse ω_{RF}:

$$\Delta z = \frac{\Delta\omega_{RF}}{\gamma G}. \tag{2.4.2}$$

Phase Encoding

After the RF pulse has been applied, a gradient in the y-direction is briefly switched on and, after a time t_y, switched off again. This gradient is applied between excitation and readout. During t_y, the spins precess at different speeds depending on their y position. After the gradient has been switched off, all spins along the y-axis return to precessing at the same speed again however, have different phases. Hence, this gradient is also called a phase encoding gradient. The accumulated spatial phase shift $\Delta\phi$(x) during the application of the phase encoding gradient G_{PE} for the time t_y is defined as

$$\Delta\phi(y) = \gamma \int_0^{t_y} G_{PE}(t) y dt. \tag{2.4.3}$$

Frequency Encoding

To distinguish between two nuclear spins which have the same y position but a different x position, a gradient varying in the x-direction is superimposed on the main magnetic field. This gradient is used for frequency encoding (or readout), since it influences the Larmor frequencies of the nuclear spins along the x-axis and completes the three dimensional spatial localisation of the MR signal. This measured MR signal by the receiver coil consists of a range of frequencies from all voxels in the slice. By applying an inverse Fourier transform to the signal, the individual frequencies can be separated.

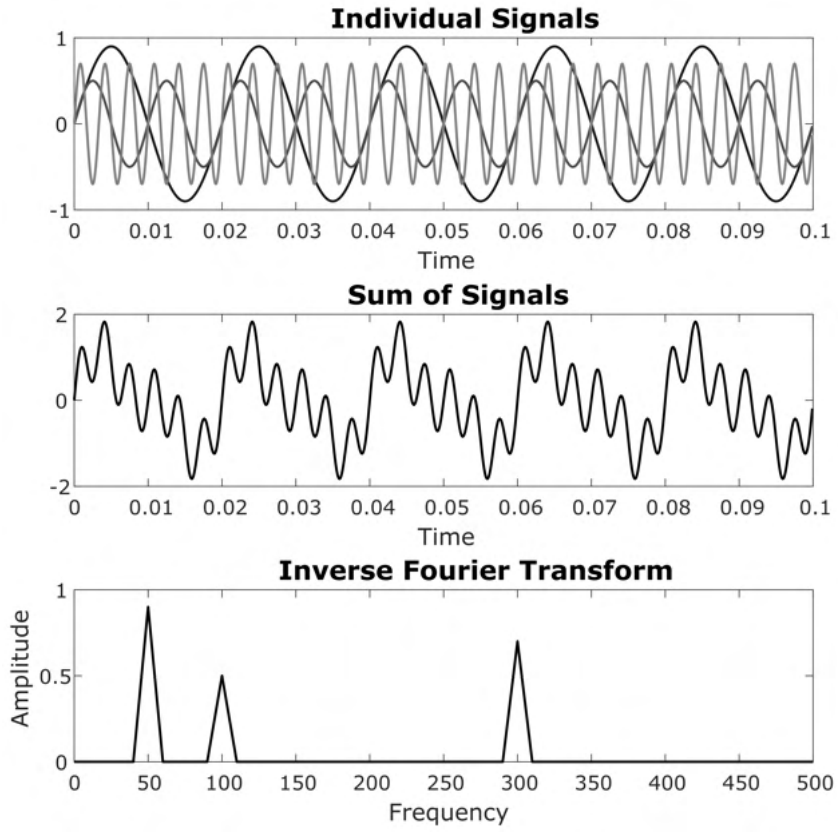

Figure 2.5: The acquired MR signal consists of different individual signals with various frequencies and amplitudes. These can be separated by the inverse Fourier transform into spectral lines with respect to their frequency. The proton density at that given location defines the amplitude of the spectral lines.

k-Space

The MR system generates the specific gradient and RF waveforms required to obtain the MR signal of any selected slice. It therefore knows the amplitudes and directions of the applied gradients from which it can determine the precise location of the raw MR signal detected in the receiver coil. This signal contains the spatial frequencies of the MR image and is stored in the so called *k*-space, where each row in *k*-space corresponds to a different phase encoding step.

The signal dS(x,y) at each point in a 2D *k*-space is defined as

$$dS(x,y) = \rho(x,y)e^{-i\gamma t(G_x x + G_y y)}dxdy \tag{2.4.4}$$

where $\rho(x,y)$ is the spin density also incorporating relaxation effects such as T_1 and T_2. Therefore, the signal over the entire slice is given by

$$S(x,y) = \int\int \rho(x,y)e^{-i\gamma t(G_x x + G_y y)}dxdy \tag{2.4.5}$$

This equation shows that S(x,y) is the Fourier transform of the spin density $\rho(x,y)$. The final MR image can be reconstructed by applying an inverse Fourier transform to the *k*-space

$$\rho(x,y) = \int\int S(x,y)e^{i\gamma t(k_x x + k_y y)}dk_x dk_y \tag{2.4.6}$$

where k_x and k_y are the *k*-space positions or spatial frequencies introduced by the applied frequency encoding ($G_x = G_{FE}$) and phase encoding ($G_y = G_{PE}$) gradients

$$k_x = \gamma \int G_x(t)dt \tag{2.4.7}$$

$$k_y = \gamma \int G_y(t)dt, \tag{2.4.8}$$

respectively. Between *k*-space and the reconstructed magnitude image, the following relationships exist:

- The spatial frequency resolutions (Δk_x, Δk_y) i.e the steps for frequency and phase encoding defines the field of view (FOV) of the magnitude image

$$FOV_x = \frac{1}{\Delta k_x}; \quad FOV_y = \frac{1}{\Delta k_y}. \tag{2.4.9}$$

- The spatial frequency field of view (FOV_k) defines the spatial resolution i.e. pixel sizes (Δx, Δy) of the magnitude image

$$\Delta x = \frac{1}{FOV_{k_x}} = \frac{1}{N_{FE} \cdot \Delta k_x}; \quad \Delta y = \frac{1}{FOV_{k_y}} = \frac{1}{N_{PE} \cdot \Delta k_y}. \tag{2.4.10}$$

  where N_{FE} and N_{PE} are the total number of frequency encoding and phase encoding steps, respectively.

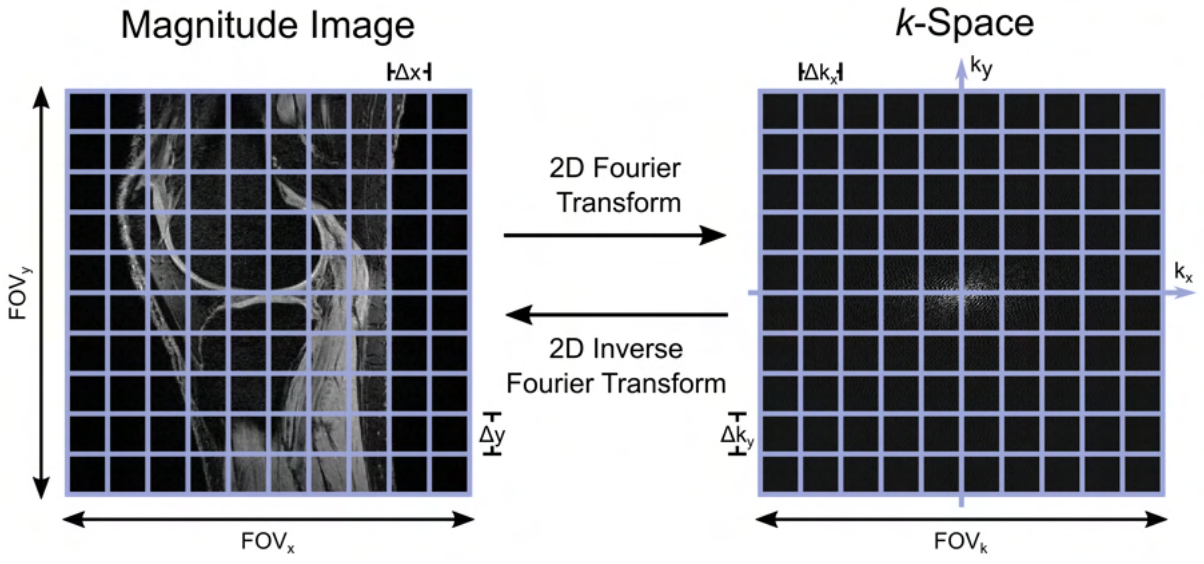

Figure 2.6: A two-dimensional inverse Fourier transform is applied to reconstruct an image from *k*-space while *k*-space is the 2D Fourier transform of the MR image.

2.4.3 Radio-Frequency Coil System

The RF coil system create the RF pulses which are used to excite the spins and tip the net magnetisation. This RF pulse created and transmitted by the RF coils is called B_1^+. While the RF transmit coil is typically a large body coil housed inside the MR system bore close to the gradient coils, the receiver coil capturing the spatially encoded MR signal is commonly placed close to the imaged volume to increase the signal-to-noise ratio (SNR) (B_1^-). The MR signal is detected by the receiver coil as a current induced by the variations in magnetic flux due to the rotation of the tipped magnetisation. Following detection, the signal is amplified, digitalised and reconstructed after processing phase and frequency data.

2.5 Basic Pulse Sequences and Quantitative Relaxation Time Mapping

In MR imaging, the FID is rarely measured directly but instead two types of echoes, the so-called spin echo and gradient echo are measured. In both cases, the magnetisation is oriented along the main magnetic field in the z-direction to begin with.

2.5.1 Single-Echo Spin Echo

In the spin echo sequence, a 90° RF excitation pulse flips the magnetisation into the transverse xy plane. Here, the spins begin to dephase due to T_2 relaxation and additional B_0 and susceptibility non-uniformities. After a time TE/2, where TE is the so-called echo time, a 180° inversion pulse is applied, which reflects the magnetisation components and accumulated phase. The spins now begin to rephase because, despite the inversion of the component distribution, the direction of rotation of the spins is retained and the slightly slower spins now precess ahead of the faster ones. After the same time TE/2, all magnetisation components point in the same direction and a clear maximum of the signal, the spin echo, is generated. However, only the dephasing effects from magnetic field inhomogeneities are rephased leaving a T_2 dependency on the generated spin echo signal.

$$S_{SE} \propto [PD] \left(1 - e^{-\frac{TR}{T_1}}\right) e^{-\frac{TE}{T_2}} \tag{2.5.1}$$

Here, PD is the proton density and TR is the repetition time, time between the first 90° excitation pulse and the next 90° pulse starting the new spin echo sequence, during which time the longitudinal magnetisation will recover. By varying TR and TE, different image contrast or weightings can be attained. Setting a long TR will result in complete longitudinal recovery, making the spin echo signal independent of tissue T_1 relaxation effects. Similarly, using a short TE will minimise the dependence of the spin echo signal on tissue T_2 relaxation effects. By minimizing the effects of both T_1 and T_2 relaxation processes through long TR and short TE, the signal will be dominated by the tissues proton density PD.

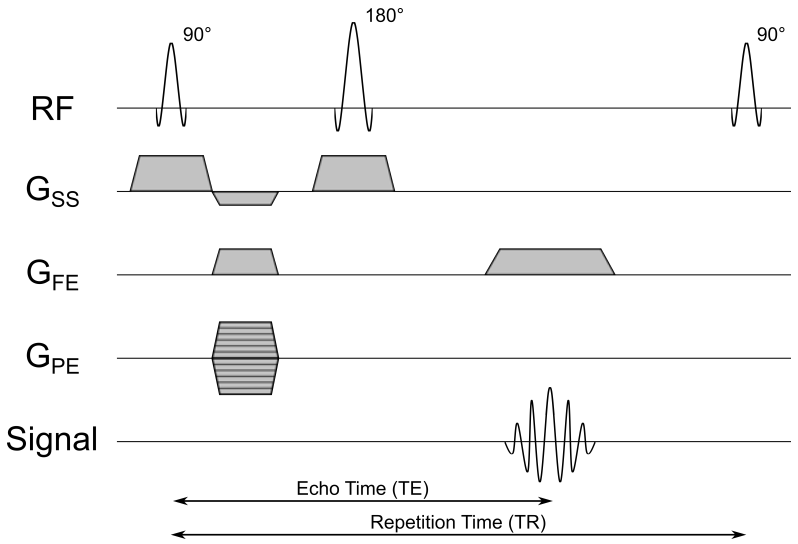

Figure 2.7: MR pulse sequence diagram for the single-echo spin echo sequence.

2.5.2 Fast Spin Echo

In contrast to the single-echo spin echo sequence, the fast spin echo (FSE) pulse sequence, also known as turbo spin echo, uses a train of 180° refocusing pulses after the initial 90° excitation pulse to generate multiple spin echoes. The total number of 180° refocusing pulses per TR is the echo train length (ETL), or turbo factor. By applying different phase encoding gradient amplitudes, each generated spin echo can be encoded as a discrete line in k-space and hence, reducing the time to acquire a full k-space. However, as with the single-echo spin echo sequence, the refocusing pulses do not recover T_2 decay effects resulting in an increased loss of signal with longer ETLs and T_2 image blurring.

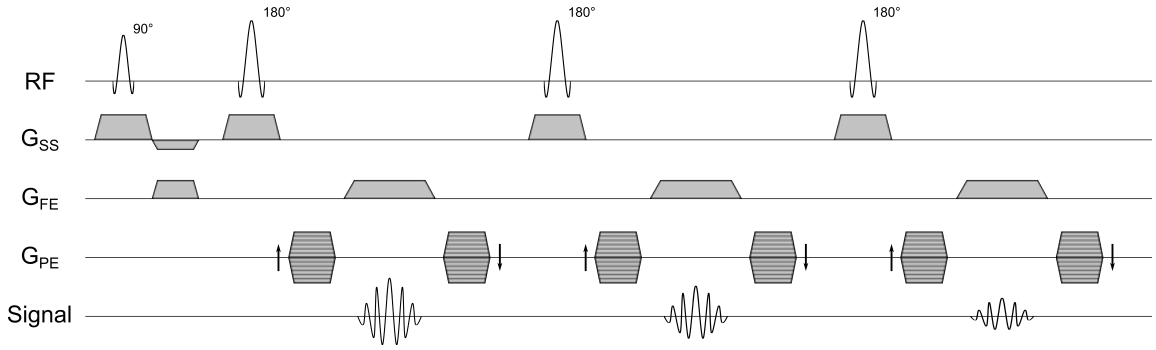

Figure 2.8: MR pulse sequence diagram for the fast spin echo sequence.

2.5.3 Gradient Echo

In the gradient echo sequence, or gradient recalled echo (GRE), the magnetisation is also first flipped into the transverse plane by an RF excitation pulse. The angle α of the initial RF excitation pulse is typically set to less than 90° to speed up acquisition by avoid saturation of longitudinal magnetisation and allowing a shorter TR. Immediately afterwards, a dephasing frequency encoding gradient is superimposed onto the main magnetic field, whereby the Larmor frequencies of the nuclear spins become location-dependent. Compared to the normal FID,

this results in an accelerated dephasing. A rephasing gradient of same amplidtude but doubled area and opposite polarity is then applied, thereby generating a gradient echo. Similar to normal FID decay, the transverse magnetisation of the gradient echo decays with T_2^* relaxation as only the additional dephasing of the spins due to the applied frequency encoding gradient is rephased while the natural dephasing due to magnetic field inhomogeneities remains.

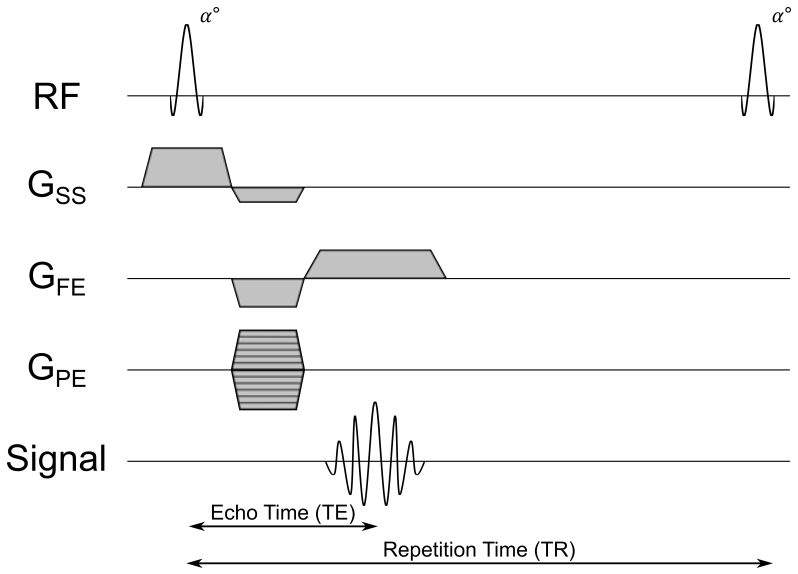

Figure 2.9: MR pulse sequence diagram for the gradient echo sequence.

2.5.4 Spoiled Gradient Echo

To allow even shorter TRs and faster gradient echo acquisitions, the spoiled gradient echo sequence, also referred to as spoiled gradient recalled echo (SPGR) on General Electric MR systems, uses RF or gradient spoiling or both to remove any residual transverse magnetisation after data acquisition. The contrast of SPGR sequences is mainly either T_1 or PD weighted, influenced by the choice of TR, TE and the flip angle α:

$$S_{SPGR} = [PD]\frac{\sin\alpha \cdot \left(1 - e^{-\frac{TR}{T_1}}\right) e^{-\frac{TE}{T_2^*}}}{1 - \cos\alpha e^{-\frac{TR}{T_1}}} \tag{2.5.2}$$

Therefore, selecting a short TR, short TE and a larger α will result in T_1-weighting, while a long TR, short TE and a small α will lead to greater PD-weighting of the image. Additionally, the SPGR signal in Equation 2.5.2 can be maximised when

$$\alpha = \alpha_E = \arccos\left(e^{-\frac{TR}{T_1}}\right) \tag{2.5.3}$$

where α_E is the Ernst angle. To note, the Ernst angle only maximises the signal of a specific tissue i.e. specific T_1.

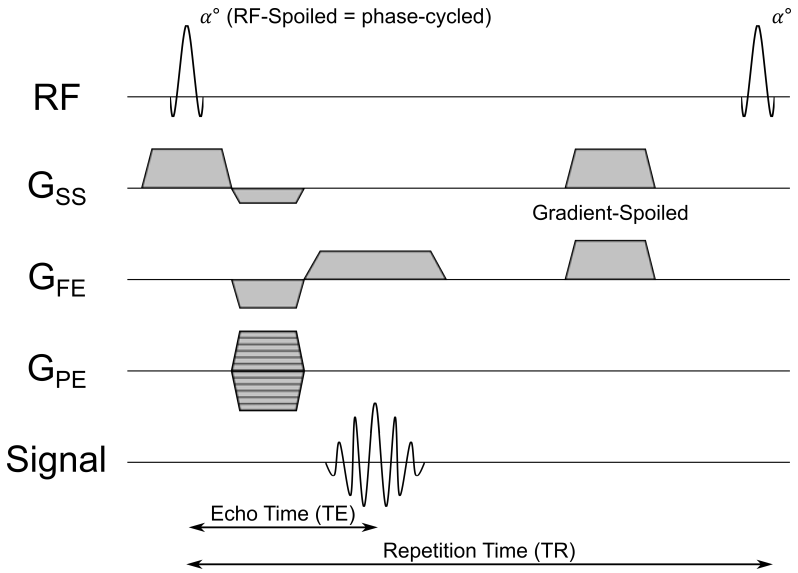

Figure 2.10: MR pulse sequence diagram for the spoiled gradient echo sequence showing both RF spoiling (phase-cycled RF excitation pulse from one TR to the next) and gradient spoiling (applied along the slice-select and readout gradient before the next TR).

2.5.5 Fat Suppression

MRI involves detecting the MR signal produced by the excitation of hydrogen atoms. The large majority of these hydrogen atoms are components of water or fat. However, in a static magnetic field, the hydrogen atoms in water and fat resonate at different Larmor frequencies ($\Delta\omega_{fw}$). This difference is also known as the chemical shift and increases with magnetic field strength. The chemical shift between water and fat is 3.5 parts per million (ppm) which corresponds to $\Delta\omega_{fw} \approx 220$ Hz at 1.5T and $\Delta\omega_{fw} \approx 440$ Hz at 3T. Additionally, fat has a short T_1 relaxation relative to other tissues with water components. Techniques used for suppressing or separating fat and water signals based on either their difference in Larmor frequencies or T_1 relaxation are commonly divided into fat saturation, selective excitation and tissue nulling inversion recovery techniques.

Fat Saturation

Fat saturation (FS) techniques typically use a narrow band 90° RF pulse at the beginning of the sequence to spectrally excite fat. A spoiler gradient is then applied shortly after to eliminate all transverse magnetisation from fat. Since there is now no longitudinal component of fat, a conventional spatially-selective RF pulse will only excite water resulting in a fat saturated image. However, this technique requires a highly homogeneous magnetic field achieve complete spectral suppression of the fat signal prior to water-only acquisition.

Selective Water Excitation

Water excitation (WE), also known as spectral-spatial fat suppression, uses two RF pulses selectively excite water signal. The first RF pulse ($B_{1,x}$) rotates both the water and fat signals to the y-axis into the transverse plane. After a time *t*, a second RF pulse ($B_{1,-y}$) is applied. During the time *t*, a phase difference of $\Delta\phi(t) = 2\pi\Delta\omega_{fw}t$ is accumulated between the water and fat magnetisations in the transverse plane. Selective water-only excitation is achieved if

the second RF pulse is applied when $\Delta\phi = 180°$ and rotates the fat-only magnetisation back towards the z-axis while leaving the water magnetisation unaffected.

Fat Nulling with Inversion Recovery

With the help of an inversion recovery sequence and exploiting the difference in T_1 relaxation time between fat and water the MR signal from fat can be suppressed. The inversion recovery sequence first inverts the longitudinal magnetisation of fat and water with a 180° RF inversion pulse. Due to the different T_1 relaxation times of water and fat the time of zero crossing (inversion time TI) is different for both. Since fat has a shorter T_1 it also has a shorter TI than other tissues with water components. By collecting data at the zero crossing of fat, the MR signal from fat can be suppressed or nulled. However, a drawback of this technique is its long acquisition time as it requires a long repetition time for the magnetisation of all tissues to return to equilibrium (M_0) before the next RF inversion pulse can be applied.

2.5.6 $T_{1\rho}$ and T_2 Quantification

In this section, methods utilised for the quantification of $T_{1\rho}$ and T_2 relaxation times are described. Quantifying these relaxation times could not only be used to characterise tissue and determine differences between healthy and diseased but also help in understanding contrast mechanisms and optimising contrast between different tissue properties [98]. The compositional MRI techniques presented here are included in the prospective study introduced in Chapter 3.

$T_{1\rho}$ Mapping

A $T_{1\rho}$-weighted image can be acquired by applying a $T_{1\rho}$ magnetisation preparation cluster prior to a typical 2D or 3D pulse sequence for data acquisition. The conventional $T_{1\rho}$ preparation cluster begins with a 90° hard pulse (tip-down) along the x-direction to flip the magnetisation into the transverse plane. This is followed by an on-resonance spin-lock RF pulse (continuous wave RF pulse) applied for a time TSL (spin-lock time) in y-direction parallel to the magnetisation and results in the transverse magnetisation to precess around it. The final pulse is a 90° hard pulse (tip-up) applied along the -x-direction and is used to flip the magnetisation back into the longitudinal plane.

This conventional spin-lock preparation cluster is thus often abbreviated as $90^\circ_x - TSL_y - 90^\circ_{-x}$. A spoiler gradient is applied immediately after the $T_{1\rho}$ preparation cluster to dephase residual transverse magnetisation [99]. In theory, during the application of the spin-lock pulse the magnetisation parallel to the pulse field direction will decay mono-exponentially with regard to TSL at a rate $1/T_{1\rho}$ and the magnetisation perpendicular to the pulse field direction also decays mono-exponentially with TSL, however at a rate $1/T_{2\rho}$ [100, 101].

The spin-lock ($T_{1\rho}$ and $T_{2\rho}$) relaxation under a spin-lock pulse applied for a time T can be described by a spin-lock relaxation matrix $E_\rho(T)$. For a spin-lock pulse applied along the y direction, $E_\rho(T)$ is defined as:

$$E_\rho(T) = \begin{bmatrix} e^{-\frac{T}{T_{2\rho}}} & 0 & 0 \\ 0 & e^{-\frac{T}{T_{1\rho}}} & 0 \\ 0 & 0 & e^{-\frac{T}{T_{2\rho}}} \end{bmatrix} = \begin{bmatrix} E_{2\rho}(T) & 0 & 0 \\ 0 & E_{1\rho}(T) & 0 \\ 0 & 0 & E_{2\rho}(T) \end{bmatrix} \tag{2.5.4}$$

A RF pulse can be represented as a rotation matrix $R_r(\Phi)$ where Φ is the pulse flip angle and r is the axis the magnetisation is rotated about. The basic rotation matrices for magnetisation rotation about x, y and z are:

$$\begin{aligned} R_x(\Phi) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\Phi) & \sin(\Phi) \\ 0 & -\sin(\Phi) & \cos(\Phi) \end{bmatrix}; \quad R_y(\Phi) = \begin{bmatrix} \cos(\Phi) & 0 & \sin(\Phi) \\ 0 & 1 & 0 \\ -\sin(\Phi) & 0 & \cos(\Phi) \end{bmatrix}; \\ R_z(\Phi) &= \begin{bmatrix} \cos(\Phi) & \sin(\Phi) & 0 \\ -\sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{2.5.5}$$

Since the duration of the 90° tip-down and tip-up pulses are much shorter than TSL, the relaxation during these are negligible [100]. With β as the flip angle of the tip-down/tip-up pulses, $\partial = 2\pi \cdot FSL \cdot TSL$ the total flip angle of the spin-locking pulse and FSL the set spin-lock frequency (typically 500 Hz), the magnetisation evolution during the conventional $90_x - TSL_y - 90_{-x}$ preparation is:

$$M(TSL) = R_{-x}(\beta) \cdot R_y(\partial) E_\rho(TSL) \cdot R_x(\beta) \cdot M(t_0) \tag{2.5.6}$$

where $M(t_0) = \begin{bmatrix} 0 & 0 & M_0 \end{bmatrix}^T$.

When inserting the respective rotation and spin-lock relaxation matrices and simplifying the resulting complex magnetisation orientation by removing the transverse components through the application of a crusher gradient, the longitudinal magnetisation can be derived as:

$$M_z = M_0 \cdot \left[E_{1\rho}(TSL) \cdot \sin^2 \beta - E_{2\rho}(TSL) \cdot \cos^2 \beta \cdot \cos \partial \right] \tag{2.5.7}$$

Thus, if the tip-down and tip-up pulses have a perfect β = 90° flip angle, the resulting longitudinal magnetisation simply follows a mono-exponential decay related to $T_{1\rho}$ relaxation:

$$M_z = M_0 \cdot E_{1\rho}(TSL) = M_0 \cdot e^{-\frac{TSL}{T_{1\rho}}} \tag{2.5.8}$$

The most common application of $T_{1\rho}$ magnetisation preparation is the voxel-wise mapping of $T_{1\rho}$ relaxation values. For this, the signal intensities from a series of differently $T_{1\rho}$-weighted images acquired with multiple spin-lock times (TSL) and constant spin-lock frequency (FSL) are fitted to the mono-exponential decay function in Equation 2.5.8.

Due to B_0 and B_1 field inhomogeneities, the anticipated spin-lock pulse direction and amplitude could be distorted. This can lead to errors in $T_{1\rho}$ quantification due to the magnetisation

being deviated from the requested spin-lock direction, resulting in a distant magnetisation evolution from the theoretically ideal mono-exponential decay[101]. Several methods have been proposed with modified $T_{1\rho}$ magnetisation preparation clusters to diminish the effects and eliminate the banding artefacts from both types of magnetic field inhomogeneities [99–104].

Rotary-Echo Spin-Lock Preparation: Given B_1 RF inhomogeneities, the flip angles of the tip-down and tip-up pulses during $T_{1\rho}$ magnetisation preparation become imperfect. The magnetisation will deviate from the intended transverse plane after the initial tip-down pulse and begin to precess at spatially varying angles around the spin-lock field during TSL [102]. To reduce the artefacts induced by the SL flip angle ∂, Charagundla et al [102] proposed using a self-compensating pulse cluster. This pulse cluster uses a 'rotary-echo' SL pulse consisting of two SL segments (SL1 and SL2) of equal duration but with opposite phase shifts: $90^{\circ}_{x} - TSL/2_{y} - TSL/2_{-y} - 90^{\circ}_{-x}$ [100, 104]. In a homogeneous B_0, the rotation angle accumulated during SL1 with a phase shift of 90° along the y direction will be ∂/2. Ideally, the accumulated rotation angle during SL2 with phase shift of -90° will equally be -∂/2 and the resulting overall longitudinal magnetisation will be independent of ∂.

$$M(TSL) = R_{-x}(\beta) \cdot R_{-y}\left(\frac{\partial}{2}\right) E_{\rho}(TSL) \cdot R_{y}\left(\frac{\partial}{2}\right) E_{\rho}(TSL) \cdot R_{x}(\beta) \cdot M(t_0) \tag{2.5.9}$$

$$M_z = M_0 \cdot \left[E_{1\rho}(TSL) \cdot \sin^2\beta - E_{2\rho}(TSL) \cdot \cos^2\beta\right] \tag{2.5.10}$$

Again, if the flip angle β = 90°, then M_z reduces to Equation 2.5.8. However, if $\beta \neq 90°$, the magnetisation is contaminated by $T_{2\rho}$ relaxation and a $T_{1\rho}$ relaxation reduced by $\sin^2\beta$.

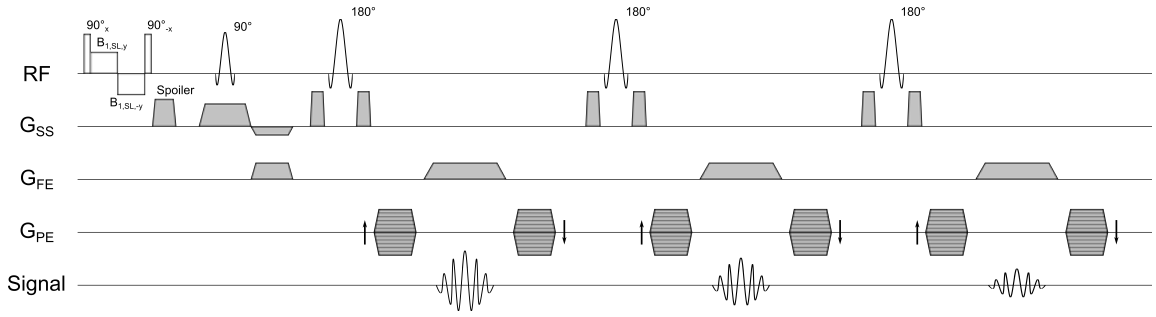

Figure 2.11: MR pulse sequence diagram for the $T_{1\rho}$-prepared sequence. The $T_{1\rho}$-preparation pulse cluster is a rotary-echo spin-lock preparation with two spin lock segments inducing opposite phase shifts. Image acquisition is performed with a 3D FSE sequence.

T_2 Mapping

Similar to $T_{1\rho}$-weighted imaging, a T_2-weighted image can be acquired by using a T_2 magnetisation preparation cluster followed by a fast imaging sequence. The cluster consists of a three-pulse sequence and is known as the driven equilibrium Fourier transform (DEFT) method [105].

The method includes a typical spin echo pulse setup followed by an additional 90° pulse to bring residual transverse magnetisation back to the longitudinal axis before T_1 relaxation

is completed. The first 90° tip-down pulse along the x-direction flips the magnetisation into the transverse plane along the y-axis. The induced transverse magnetisation decays because of spin-spin relaxation (T_2) and of magnetic field inhomogeneity (T_2^*). After a time τ, a 180° pulse is applied along the x-direction to reverse the dephasing of the transverse magnetisation due to field inhomogeneities, which refocuses the magnetisation and produces an echo at t = 2τ. The last 90° tip-up pulse applied at the peak of the echo along the x-direction restores the magnetisation along the z-axis [106]. The T_2 preparation cluster can be abbreviated as $90^\circ_x - \tau - 180^\circ_x - \tau - 90^\circ_x$ with τ being the time between the pulses. A spoiler gradient follows the T_2 preparation cluster to eliminate any unwanted transverse magnetisation caused by pulse imperfections. Therefore, the magnitude of longitudinal magnetisation M_z after the final 90° tip-up pulse depends on the extent of T_2 spin-spin relaxation during the time interval 2τ = TE:

$$M_z = M_0 \cdot e^{-\frac{TE}{T2}} \tag{2.5.11}$$

Quantitative T_2 maps can be calculated by fitting the signal intensities of acquired data sets with different echo times TE on a voxel-by-voxel basis to the mono-exponential decay function in Equation 2.5.11.

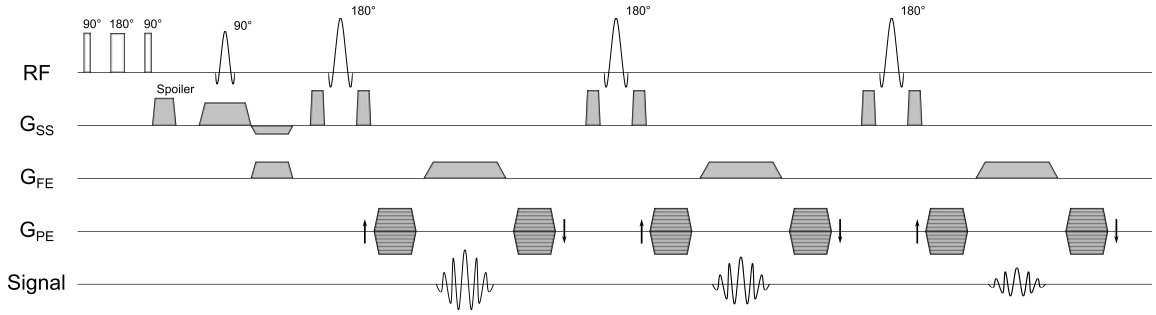

Figure 2.12: MR pulse sequence diagram for the T_2-prepared sequence. The T_2-preparation pulse cluster is a driven equilibrium preparation ($90^\circ_x - \tau - 180^\circ_x - \tau - 90^\circ_x$). Image acquisition is performed with a 3D FSE sequence.

Chapter 3

Effectively Measuring Exercise-related Variations in $T_{1\rho}$ and T_2 Relaxation Times of Healthy Articular Cartilage

Published as *Original Research* in *Journal of Magnetic Resonance Imaging*, 52(6):1753-1764; December 2020.

3.1 Introduction

Over the last two decades in vivo magnetic resonance imaging (MRI) has increasingly been used to determine the mechanical properties of knee articular cartilage. Previous studies have shown that cartilage loading activities affect the morphology and biochemical composition of articular cartilage and have provided important information on the behaviour of cartilage when exposed to different compressive loads [54, 71, 107]. $T_{1\rho}$ and T_2 relaxation time mapping techniques allow the assessment of cartilage compositional alterations in response to joint loading as they have been demonstrated to be sensitive to variations in the water and macromolecular content of cartilage [39, 42, 108]). Normalised changes in $T_{1\rho}$ and T_2 relaxation times of cartilage following different exercise regimes have been shown to be in the order of -2.6% to -14.3% and +3.7% to -12.5%, respectively [62, 71, 107, 109, 110]. Since the measured changes resulting from joint loading can be small, determining the intra-sessional repeatability of these quantitative measures is essential for reliable assessment of joint loading-related effects on cartilage structure and composition.

A systematic review showed that studies assessing the repeatability of these quantitative relaxation techniques without any joint loading activity have reported root-mean-squared coefficient of variation (RMS-CV) for large regional analysis of $T_{1\rho}$ values in the range of 2.3% – 6.3% and of T_2 values in the range of 2.3% – 6.5% [41]. When sub-regional or laminar cartilage analysis was performed, test-retest CVs for $T_{1\rho}$ were up to 19% and for T_2 as high as 22% [41]. Intra-sessional repeatability assesses the repeatability of measurements of i) consecutive scans without repositioning and ii) consecutive scans with repositioning of the subject [111]. Evaluating the repeatability of consecutive scans without repositioning is important when measuring $T_{1\rho}$ and T_2 at multiple time-points after joint-loading for determining longitudinal cartilage recovery as previous studies have reported [54, 60, 112].

Healthy cartilage is maintained with regular deformation and compression of the cartilage

structure and its extracellular matrix (ECM) through physiological loading, such as experienced during exercise [11, 13]. However, both overuse and disuse can have degenerative effects on the cartilage and are important risk factors in the development of osteoarthritis (OA) [13, 16, 21]. When exposing the cartilage repeatedly to excessive loads, such as may occur during high-impact sports or, to minimal or no load following injury, the cartilage structure and microstructure begin to break down [13, 22]. Morphological changes in articular cartilage volume, thickness and joint space narrowing are not necessarily present in the early stages of OA and may change very slowly during disease progression. Therefore, measuring differences in cartilage deformational responses during or after loading may represent a more sensitive biomarker for detecting the early onset of OA [23, 113].

The aim of this study was to measure the intra-sessional repeatability of both $T_{1\rho}$ and T_2 of knee articular cartilage and to determine if these quantitative relaxation measurement techniques are sensitive to permit effective measurement of short-term cartilage compositional responses after a joint loading activity.

3.2 Methods

All imaging was performed on a 3 T MRI system (MR750, GE Healthcare, Waukesha, WI, USA) using an 8-channel transmit/receive knee coil (Invivo, Gainesville, FL, USA). Participant imaging had local ethical approval, and written informed consent was provided by each participant.

3.2.1 Study Procedures

Phantom Repeatability

To assess the test-retest repeatability of the quantitative $T_{1\rho}$ and T_2 relaxation time measurements for a range of relaxation times, two consecutive $T_{1\rho}$ and T_2 relaxation mapping datasets were obtained from a phantom. The phantom consisted of five vials having different $T_{1\rho}$ and T_2 relaxations. Two vials had $T_{1\rho}$ and T_2 relaxation times similar to cartilage (40 - 50 ms) at 3 T while the relaxation times of the remaining three vials were greater [114, 115]. To additionally assess the inter-sessional variability (scanning the same phantom on different days), two further $T_{1\rho}$ and T_2 relaxation mapping datasets were acquired two days later. On each day, the same knee coil and setup was used with the phantom centred in the coil.

Group 1: In Vivo Repeatability Study

To assess the intra-sessional repeatability of $T_{1\rho}$- and T_2-relaxation mapping of cartilage, the right knee of ten healthy participants (five men, five women, mean age 28.9 ± 5.5 years) with no current knee pain symptoms, nor known history of joint disorder was imaged. Imaged knees were unloaded for 15 minutes prior to the imaging session to minimise short-term loading effects on the joint.

The MR session consisted of a sagittal 3D fat-saturated spoiled gradient recalled-echo (3D-FS SPGR) sequence, and sagittal $T_{1\rho}$- and T_2-mapping sequences. For details on pulse sequence parameters used, see section 'Sequence Parameters' below. Following repositioning of the participant and imaged knee, two consecutive acquisitions of $T_{1\rho}$- and T_2-mapping were

performed using the same pulse sequences as before repositioning (Figure 3.1a). During knee repositioning, the participants removed their knee from the coil and sat up on the side of the MR table. The coil was repositioned, followed by participant positioning. The time required for repositioning and the continuation of the imaging protocol was approximately five minutes.

Group 2: Exercise and Recovery Study

A second group were used to assess the magnitude of effect that mild exercise has on $T_{1\rho}$- and T_2-relaxation mapping of cartilage. The right knee of nine healthy participants (five men, four women, mean age 31.6 ± 6.0 years) with no current knee pain symptoms, nor known history of joint disorder was imaged. Imaged knees were unloaded for 15 minutes prior to the imaging session to minimise short-term loading effects on the joint.

The study design consisted of a 3D- FS SPGR sequence, followed by $T_{1\rho}$- and T_2-relaxation imaging before exercise, and at four time-points after exercise to assess cartilage compositional recovery. The standardised exercise protocol involved five minutes of stepping onto a step-stool (height $\approx$ 24cm) with one leg and stepping down onto the other side of the step-stool with the leg to be imaged (Figure 3.1b). This resulted in approximately 20 stepping cycles per minute in which the knee joint was repeatedly loaded.

The first post-exercise $T_{1\rho}$- and T_2-mapping sequences were acquired approximately at five and ten minutes after patient positioning, respectively. The post-exercise imaging protocol took approximately 45 minutes.

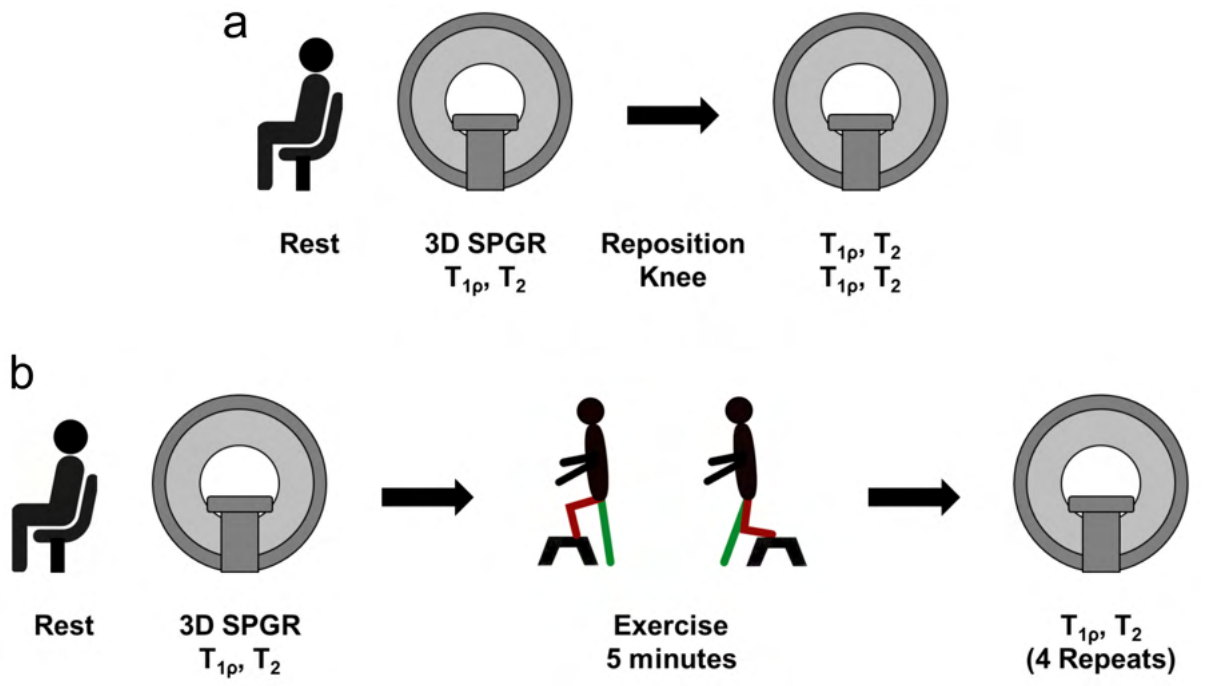

Figure 3.1: Summary of MR sessions performed. **a**: In vivo assessment of intra-sessional repeatability of cartilage $T_{1\rho}$ and T_2 mapping. After having the participant sit and keep the imaged knee in an unloaded state for approximately 15 minutes prior to imaging, initial $T_{1\rho}$ and T_2 relaxation mapping was acquired. Following knee repositioning, two successive $T_{1\rho}$ and T_2 relaxation mapping measurements were acquired. **b**: In vivo assessment of the change in cartilage composition following mild exercise. The imaged knee (green) was kept in an unloaded state for approximately 15 minutes before acquiring the initial $T_{1\rho}$ and T_2 relaxation measurements. Following mild exercise, four repeats of $T_{1\rho}$ and T_2 relaxation mapping measurements were acquired to evaluate cartilage compositional change and recovery following exercise.

3.2.2 Sequence Parameters

3D-FS SPGR

The sagittal 3D-FS SPGR sequence parameters were: acquisition time = 6:52 min; field-of-view = 150 x 128 x 136 mm^3, matrix size=512 x 380 x 136 zero-fill interpolated to 512 x 512 x 136, reconstructed voxel size = 0.29 x 0.29 x 1.00 mm^3, TR = 25.8 ms, TE = 6.8 ms, flip angle = 25°, coil acceleration factor (ASSET) = 2, number of excitations (NEX) = 0.7, bandwidth = ±11.9 kHz, with chemical shift selective fat-suppression.

$T_{1\rho}$ Mapping

$T_{1\rho}$ maps were obtained with a sagittal $T_{1\rho}$-prepared pseudo-steady-state 3D fast spin echo (PSS 3D-FSE) sequence using a rotary-echo spin-lock preparation to minimise B_1 non-uniformity effects [102, 116]. Images were acquired using the following parameters: acquisition time = 5:23 min; matrix = 320 x 256 interpolated to 512 x 512; FOV = 160 x 144 mm^2; reconstructed voxel size = 0.31 x 0.31 x 3.00 mm^3; flip angle = 90°; TR = 1580 ms; spin lock time (TSL) = 1, 10, 20, 35 ms; 72 slices per TSL; echo train length = 45; NEX = 0.5; and bandwidth = ±62.5 kHz. The $T_{1\rho}$ maps were created using a log-linearised least-squares algorithm to fit a mono-exponential decay function to the signal intensities

$$S = S_0 \cdot e^{-\frac{TSL}{T_{1\rho}}}. \tag{3.2.1}$$

Where $S(TSL)$ is the signal intensity of the $T_{1\rho}$-weighted image at a specific TSL and S_0 is the initial magnetisation / signal intensity. $T_{1\rho}$ relaxation times > 130ms in $T_{1\rho}$ maps were excluded from analysis to avoid partial volume effects with synovial fluid [82, 117].

T_2 Mapping

T_2 maps were obtained with a sagittal T_2-prepared PSS 3D-FSE sequence using a composite 90x - 180y - 90x pulse train for T_2-preparation [116, 118]. Images were acquired using the following parameters: acquisition time = 5:25 min; matrix = 320 x 256 interpolated to 512 x 512; FOV = 160 x 144 mm^2; reconstructed voxel size = 0.31 x 0.31 x 3.00 mm^3; flip angle = 90°; TR = 1580 ms; TEs = 6.5, 13.4, 27.0, 40.7 ms; 72 slices per TE; echo train length = 45; NEX = 0.5; and bandwidth = ±62.5 kHz. The T_2 maps were created using a log-linearised least-squares algorithm to fit a mono-exponential decay function to the signal intensities

$$S = S_0 \cdot e^{-\frac{TE}{T_2}}. \tag{3.2.2}$$

where $S(TE)$ is the signal intensity of the T_2-weighted image at a specific TE and S_0 is the initial magnetisation / signal intensity. As with $T_{1\rho}$, T_2 relaxation times > 100ms in T_2 maps were excluded from analysis to avoid partial volume effects with synovial fluid [82, 117].

3.2.3 Imaging Analysis

Phantom Repeatability

Mean relaxation times from all five vials of the phantoms were determined using rectangular regions-of-interest (ROIs) placed on two central sequential slices of the sagittal $T_{1\rho}$ and T_2 maps.

In Vivo Surface Analysis

All $T_{1\rho}$- and T_2-weighted images were rigidly registered to the high-resolution 3D-FS SPGR images using the Elastix toolbox [119] before calculating the respective quantitative maps.

Surface-based analysis (3D Cartilage Surface Mapping, 3D-CaSM) of femoral, tibial and patellar cartilage was performed using the freely available Stradwin software version 5.4a (University of Cambridge Department of Engineering, Cambridge, UK, now freely available as 'StradView' at http://mi.eng.cam.ac.uk/Main/StradView/) [3]. After creating sparse manual cross-sections (on every 2nd – 4th sagittal slice) of the patella, tibia, and femur including their surrounding cartilage on the 3D-FS SPGR datasets, a triangulated surface mesh object of each segmented bone-cartilage structure was automatically generated using shape-based interpolation and the regularised marching tetrahedra method [120]. Following cartilage thickness calculation and the generation of inner and outer cartilage surfaces, these surfaces were used to analyse the registered quantitative $T_{1\rho}$ and T_2 maps. At each vertex, the $T_{1\rho}$ and T_2 values along a perpendicular line between inner and outer surface (surface normal) were sampled and averaged.

Canonical (average) femoral, tibial and patellar meshes were created from all participants to be able to compare the $T_{1\rho}$ and T_2 value distributions between participants. Canonical surfaces were calculated from all participants involved in the exercise and recovery imaging. All quantitative surface data from both the repeatability and exercise-recovery cohorts were mapped onto the canonical surface following surface registration. Canonical surface generation and the subsequent registration and mapping of the individual surfaces was performed using the freely available wxRegSurf software version 18 (University of Cambridge Department of Engineering, Cambridge, UK, freely available at http://mi.eng.cam.ac.uk/ ahg/wxRegSurf/) [121]. The full 3D-CaSM analysis pipeline is illustrated in Figure 3.2.

3.2.4 Statistical Analysis

Phantom Repeatability

Coefficients of variation (CVs) were calculated from the two successive repeatability scans on each day ($CV_{Phant,Day1}$, $CV_{Phant,Day2}$) for all five vials using

$$CV = \frac{\sigma}{\mu} \tag{3.2.3}$$

with σ being the within-vial standard deviation and μ the within-vial mean of measurements. The intra-phantom variability was evaluated by calculating the CV from the mean and standard deviation of the relaxation values obtained from both days ($CV_{Phant,All}$).

Group 1: In Vivo Repeatability Study

The intra-sessional repeatability of $T_{1\rho}$ and T_2 acquisitions was assessed by calculating root-mean-square average coefficients of variation (RMS-CV) from the surface-averaged $T_{1\rho}$ and T_2 measurements of all participants for femoral, medial tibial, lateral tibial and patellar cartilage surfaces. The RMS-CV between repeatability measurements 1 (before repositioning) and 2 (first measurement following repositioning) were calculated (RMS-CV_{S1-S2}) to evaluate the effects of knee repositioning on repeatability. The RMS-CV between measurements 2 and 3 (with no repositioning between both measurements) were determined to assess repeatability without knee repositioning (RMS-CV_{S2-S3}).

The smallest detectable difference (SDD) [122] was calculated as the repeatability coefficient from the ±95% confidence intervals from a Bland-Altman analysis [123] of all surface vertices of the repeatability data for all four cartilage surfaces and for both $T_{1\rho}$ and T_2.

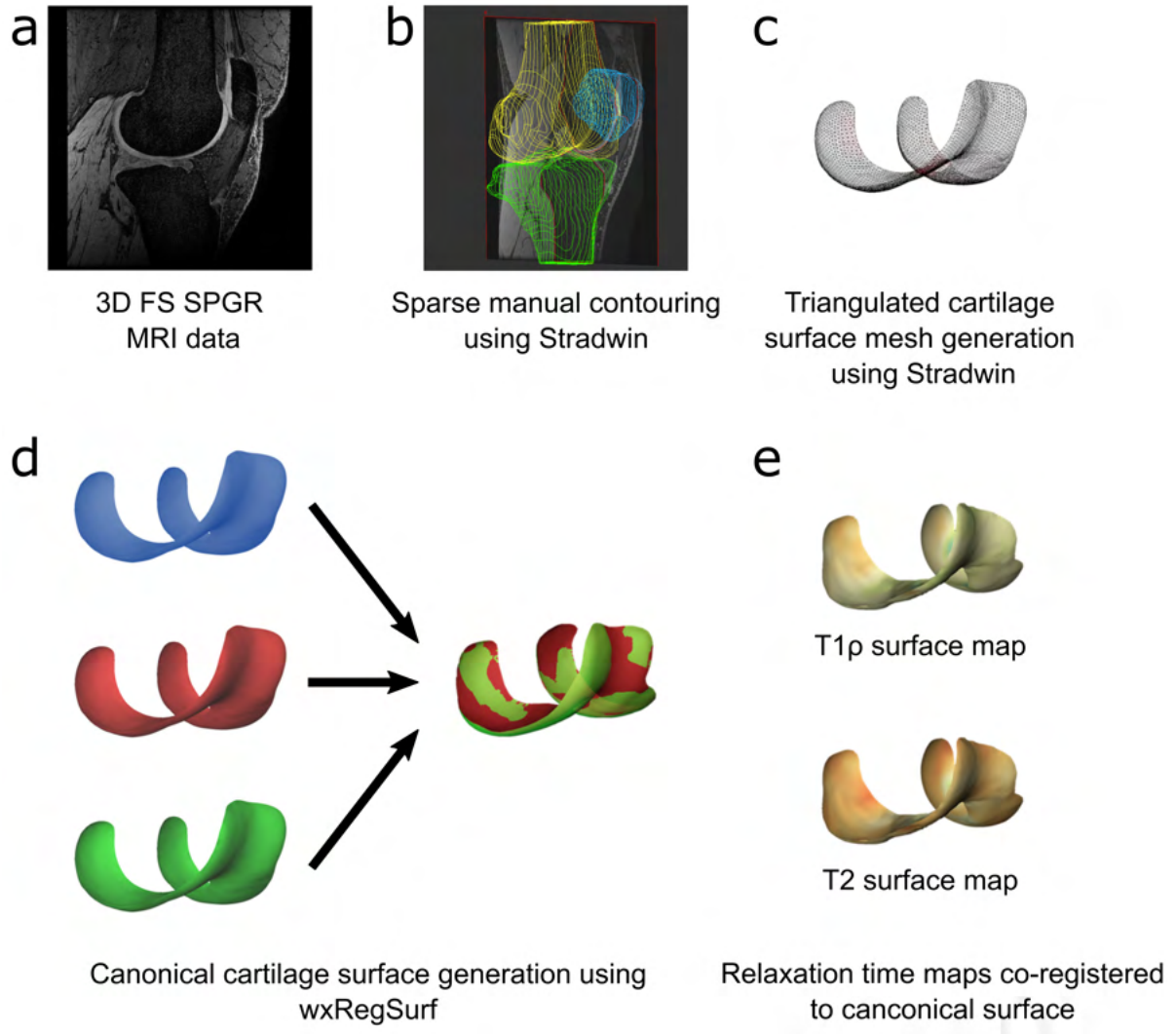

Figure 3.2: Summary of 3D-CaSM analysis pipeline illustrated for femoral cartilage surface. The 3D-FS SPGR datasets (**a**) were used to creating sparse manual contouring (on every 2nd – 4th sagittal slice) of the patella (blue), tibia (green), and femur (yellow) including their surrounding cartilage (**b**). Following the generation of unique triangulated surface mesh objects of each cartilage surface (**c**) and for each participant, canonical cartilage surfaces were calculated (**d**). All the quantitative surface data ($T_{1\rho}$ and T_2) from both the repeatability and exercise-recovery groups were mapped onto the canonical surface following surface registration (**e**).

Group 2: Exercise and Recovery Study

To determine the effects of the dynamic joint-loading stepper activity on mean MR relaxation times of entire cartilage surfaces, linear mixed-effects models with timepoint as a fixed effect and participant as a random effect for each surface/parameter combination were created. For all statistical analysis, a level of significance of 0.05 was used.

The upper (+1.96 $\cdot$ σ) and lower (-1.96 $\cdot$ σ) limits of agreement as determined from the ±95% confidence intervals of the Bland-Altman plots of the repeatability data were used to establish thresholds.

Exercise-induced changes in vertex-wise $T_{1\rho}$ and T_2 relaxation times greater than the SDD signify variations which have a 95% probability of representing a true change rather than a variation due to measurement error [124]. Thresholds were determined for all four cartilage surfaces of interest. The determined thresholds were applied to the canonical surface data to only present cartilage regions undergoing a statistically significant exercise-induced compositional change at each surface vertex.

Vertex-wise percentage changes in $T_{1\rho}$ ($\%T_{1\rho}$ change) and T_2 ($\%T_2$ change) following exercise were calculated as the normalised change in cartilage relaxation time measurements

$$\%T_{relax} = 100 \cdot \frac{T_{relax,post} - T_{relax,pre}}{T_{relax,pre}} \tag{3.2.4}$$

where $T_{relax,post}$ is the relaxation time measurement at a post exercise timepoint and $T_{relax,pre}$ is the relaxation time measurement prior to exercise.

The variability of $T_{1\rho}$ and T_2 relaxation values during cartilage compositional recovery following exposure to the mild stepping exercise was assessed only in the cartilage regions determined as regions experiencing significant exercise responses.

3.3 Results

3.3.1 Phantom Imaging

The phantom test-retest repeatability on both days ($CV_{Phant,Day1}$, $CV_{Phant,Day2}$) was $\leq$2.29% for $T_{1\rho}$ and $\leq$0.74% for T_2 relaxation time measurements for all five vials. The CVs for the two phantoms having relaxation times comparable to cartilage were $\leq$0.64% for $T_{1\rho}$ and $\leq$0.21% for T_2. The inter-sessional repeatability ($CV_{Phant,All}$) calculated from all phantom repeatability scans over both days was $\leq$2.94% and $\leq$1.43% for $T_{1\rho}$ and T_2 relaxation time measurements, respectively. The measured relaxation times and determined CVs are listed in Table 3.1.

3.3.2 Group 1: In Vivo Repeatability Study

The intra-sessional repeatability RMS-CV for in vivo relaxation time measurements averaged over the entire femoral, medial tibial, lateral tibial and patellar cartilage surfaces are listed in Table 3.2. The determined mean ± standard deviation (SD) of $T_{1\rho}$ relaxation times of repeatability scan 1 from all participants in group 1 for femoral, lateral tibial, medial tibial and patellar cartilage surfaces were 50.1 ± 2.6 ms, 44.0 ± 3.3 ms, 44.0 ± 4.0 ms and 51.2 ± 3.5 ms. Mean ±

Table 3.1: Coefficients of variation (CV) for phantom $T_{1\rho}$ and T_2 repeatability measurements. The phantom consisted of five vials with different $T_{1\rho}$ and T_2 relaxation times. The phantom was scanned on two days with two successive $T_{1\rho}$ and T_2 measurements on each day ($CV_{Phant,Day1}$, $CV_{Phant,Day2}$). The inter-sessional CV ($CV_{Phant,All}$) was calculated from all measurements of both days.

Parameter	Vial	Day 1		Day 2		$CV_{Phant,ALL}$ [%]
		Mean (SD) [ms]	$CV_{Phant,Day1}$ [%]	Mean (SD) [ms]	$CV_{Phant,Day2}$ [%]	
$T_{1\rho}$	1	43.96 (0.02)	0.05	44.46 (0.28)	0.64	0.75
	2	47.41 (0.20)	0.42	47.75 (0.16)	0.3	0.52
	3	133.33 (1.13)	0.85	138.76 (3.17)	2.29	2.71
	4	157.56 (1.32)	0.84	164.27 (3.44)	2.09	2.75
	5	214.54 (4.29)	2.00	204.79 (0.74)	0.36	2.94
T_2	1	41.11 (0.01)	0.04	41.02 (0.08)	0.21	0.18
	2	45.40 (0.09)	0.19	45.24 (0.02)	0.05	0.24
	3	130.32 (0.02)	0.01	133.52 (0.58)	0.43	1.43
	4	144.63 (0.14)	0.10	144.01 (1.06)	0.74	0.49
	5	179.31 (0.38)	0.21	181.88 (0.59)	0.32	0.85

SD of T_2 relaxation times for femoral, lateral tibial, medial tibial and patellar cartilage surfaces were 37.2 ± 1.6 ms, 32.0 ± 1.5 ms, 32.0 ± 2.3 ms and 35.5 ± 2.9 ms.

Table 3.2: Root-mean-squared coefficients of variation (RMS-CV) for in vivo $T_{1\rho}$ and T_2 repeatability measurements. For RMS-CV calculation, the vertex-wise $T_{1\rho}$ and T_2 measurements were averaged over whole femoral, lateral tibial, medial tibial and patellar cartilage surfaces. Between repeatability scans 1 and 2, the knee was repositioned ($RMS\text{-}CV_{S1-S2}$). Repeatability scans 2 and 3 were obtained successively and without knee repositioning ($RMS\text{-}CV_{S2-S3}$).

Cartilage	$T_{1\rho}$		T_2	
Surface	$RMS\text{-}CV_{S1-S2}$ [%]	$RMS\text{-}CV_{S2-S3}$ [%]	$RMS\text{-}CV_{S1-S2}$ [%]	$RMS\text{-}CV_{S2-S3}$ [%]
Femoral	0.15	0.24	0.99	0.10
Lateral Tibial	0.26	0.03	2.03	0.30
Medial Tibial	0.41	0.90	1.37	1.09
Patellar	4.81	0.05	1.39	0.22

Knee repositioning showed the greatest effect on the mean surfaced-averaged $T_{1\rho}$ relaxation time values of the patellar cartilage (51.2 ms → 54.8 ms, $RMS\text{-}CV_{S1-S2}$ = 4.8%) and the mean surfaced-averaged T_2 relaxation times of the lateral tibial cartilage (32.0 ms → 32.9 ms, $RMS\text{-}CV_{S1-S2}$ = 2.0%).

The Bland-Altman plots for vertex-wise $T_{1\rho}$ and T_2 repeatability measurements with knee repositioning of all four cartilage surfaces under investigation are shown in Figure 3.3a and Figure 3.3b, respectively.

The determined SDD and 95% limits of agreement from the Bland-Altman plots of all four cartilage surfaces and both compositional MRI methods are listed in Table 3.3.

3.3.3 Group 2: Exercise and Recovery Study

The $T_{1\rho}$ and T_2 relaxation times averaged over whole femoral, lateral tibial, medial tibial and patellar cartilage surfaces are illustrated in Figure 3.4. The determined mean baseline $T_{1\rho}$

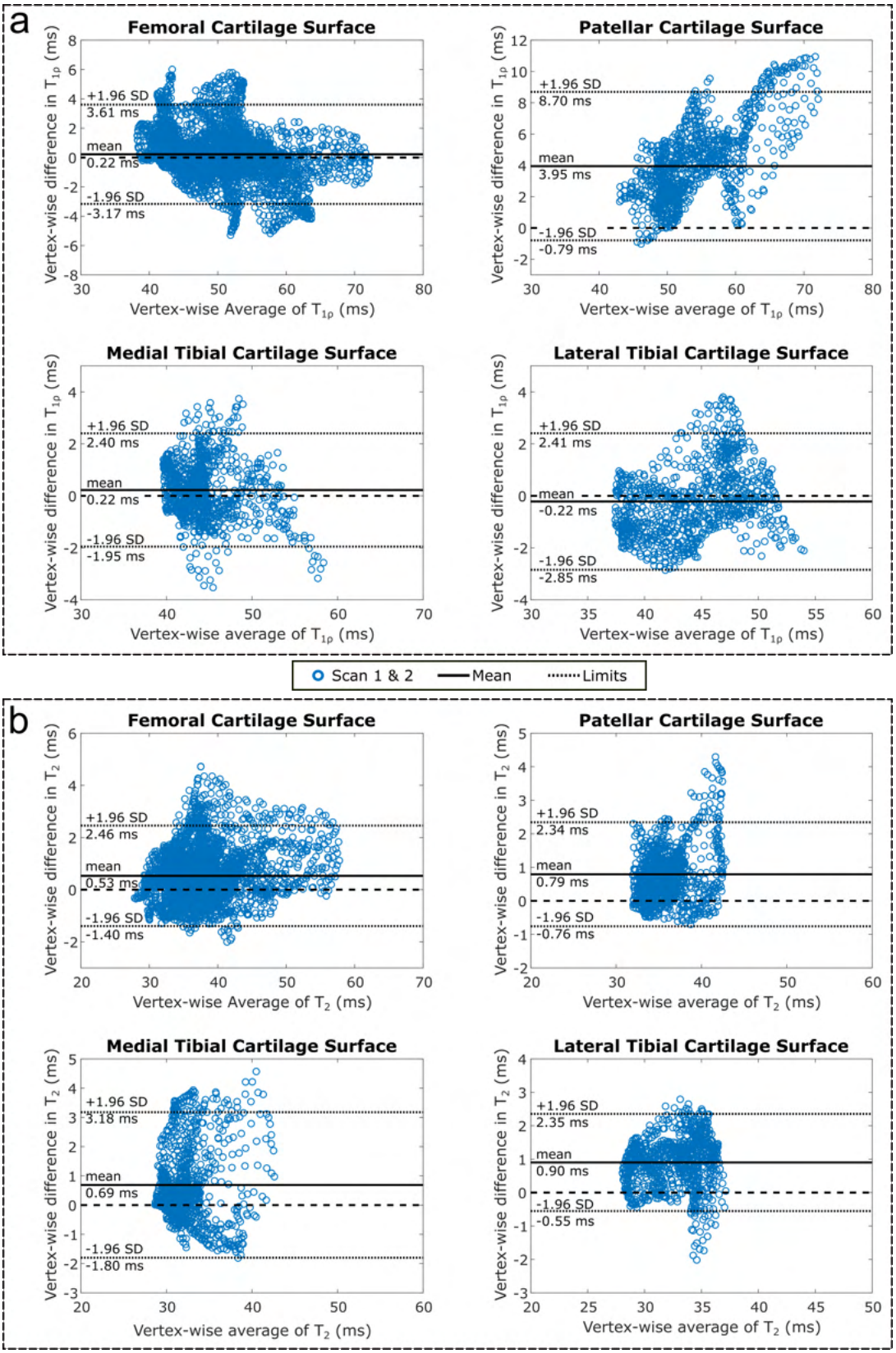

Figure 3.3: a: Bland-Altman plots showing the difference in $T_{1\rho}$ measurements with knee repositioning between repeatability acquisition 1 and 2 (blue circles) against their mean values. **b**: Bland-Altman plots showing the difference in T_2 measurements with knee repositioning between repeatability acquisition 1 and 2 against their mean values. The dotted lines represent the 95% limits of agreement; the solid line is the overall mean difference from all difference measurements.

Table 3.3: Determined smallest detectable differences and ±95% limits of agreement from Bland-Altman analysis for both $T_{1\rho}$ and T_2 and for all cartilage surfaces. Abbreviations: LoA, limits of agreement; SDD, smallest detectable difference.

Cartilage	$T_{1\rho}$		T_2	
Surface	SDD [ms]	+/- 95% LoA [ms]	SDD [ms]	+/- 95% LoA [ms]
Femoral	3.4	+3.6 / -3.2	1.9	+2.5 / -1.4
Lateral Tibial	2.6	+2.4 / -2.9	1.5	+2.4 / -0.6
Medial Tibial	2.2	+2.4 / -2.0	2.5	+3.2 / -1.8
Patellar	4.8	+8.7 / -0.8	1.6	+2.3 / -0.8

relaxation times from the exercise-recovery cohort for femoral, lateral tibial, medial tibial and patellar cartilage surfaces were 50.9 ± 3.6 ms, 44.3 ± 4.5 ms, 44.9 ± 3.7 ms and 51.2 ± 8.9 ms. Mean baseline T_2 relaxation times for femoral, lateral tibial, medial tibial and patellar cartilage surfaces were 38.0 ± 2.0 ms, 34.4 ± 2.3 ms, 32.9 ± 3.0 ms and 34.6 ± 4.2 ms. There was a statistically significant group-averaged change of T_2 of the lateral tibia over time (b [95% CI] = -0.43 [-0.83, -0.04], $p < 0.05$). No other surface/parameter combination demonstrated a statistically significant change over time at the group level. There was significant variation in change over time between participants for medial tibial $T_{1\rho}$ (SD [95% CI] = 1.04 [0.62,1.75], $p < 0.05$). The results of the linear mixed-effects models for each region are provided in Table 3.4.

Table 3.4: Results of the linear mixed-effects models for each region with timepoint as a fixed effect and participant as a random effect for each surface/parameter combination. *Variance in intercepts across participants was highly significant ($p < 0.001$) in all cases. Only variance of slopes is presented here.

Parameter	Surface	Fixed Effect (Timepoint)		Random Effect (Participant)*	
		b (95% CI)	p	Slope SD (95% CI)	p
$T_{1\rho}$	Femur	-0.07 (-0.46, 0.31)	0.71	0.28 (0.03, 2.68)	0.76
	Medial Tibia	0.34 (-0.42, 1.10)	0.39	1.04 (0.62, 1.75)	0.006
	Lateral Tibia	-0.15 (-0.71, 0.40)	0.59	0.48 (0.15, 1.55)	0.12
	Patella	-0.15 (-0.74,0.45)	0.62	0.27 (0.01, 5.41)	0.78
T_2	Femur	-0.15 (-0.38, 0.07)	0.19	0.27 (0.13, 0.56)	0.47
	Medial Tibia	0.14 (-0.24, 0.53)	0.69	0.17 (0, 68.38)	0.99
	Lateral Tibia	-0.43 (-0.82, -0.04)	0.04	0.36 (0.12, 1.02)	0.28
	Patella	-0.11 (-0.48, 0.26)	0.55	0.28 (0.08, 0.94)	0.13

Figure 3.5 and Figure 3.7 highlight the cartilage regions experiencing statistically significant changes in $T_{1\rho}$ and T_2 relaxation times following the mild stepping exercise, respectively. Correspondingly, Figure 3.6 and Figure 3.8 illustrate the alteration ('recovery') in participant-averaged femoral, patellar, medial and lateral cartilage $T_{1\rho}$ and T_2 percentage ($\%T_{1\rho}$ and $\%T_2$) changes determined from the four post-exercise measurements (scans 2 – 5) and the one pre-exercise baseline measurement (scan 1).

Table 3.5 shows the total number of vertices of each canonical cartilage surface and the percentage of cartilage surface area covered in regions experiencing changes (increases and

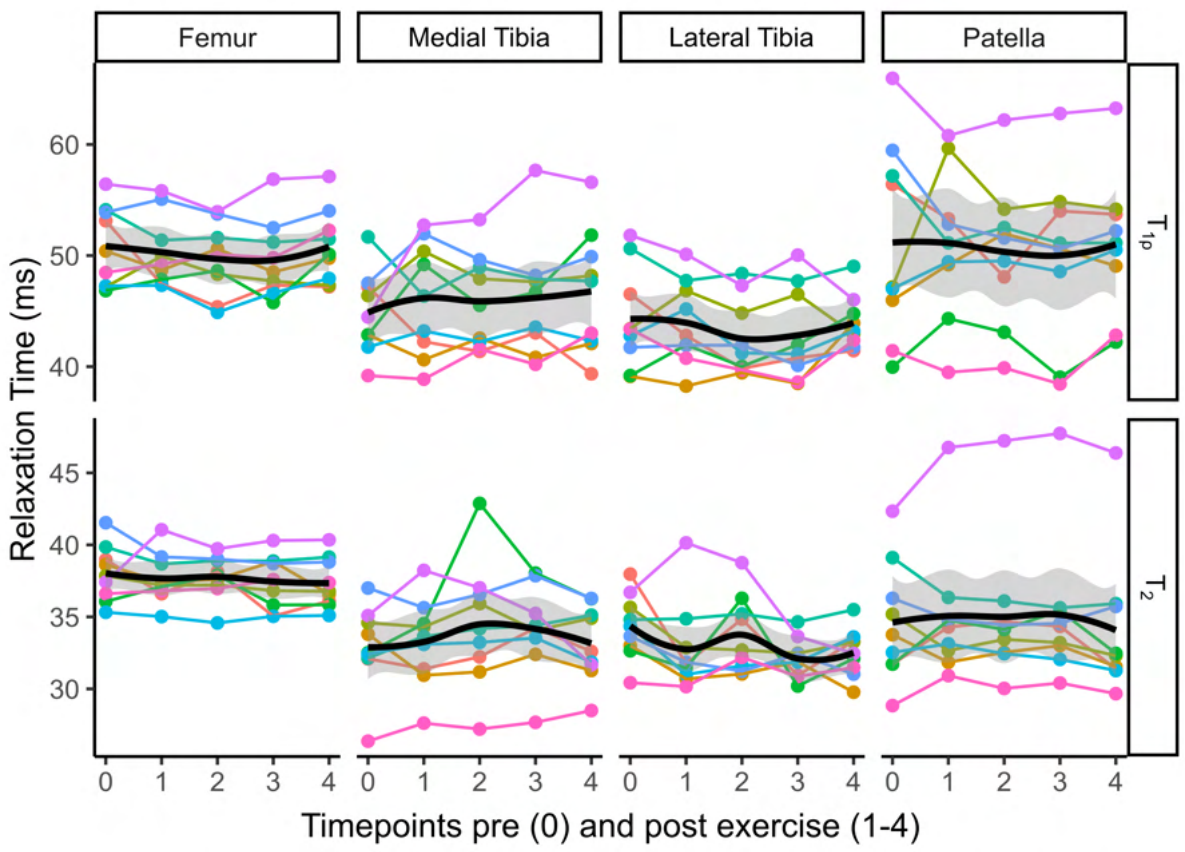

Figure 3.4: $T_{1\rho}$ (top) and T_2 measurements (bottom) averaged over whole femoral, medial tibial, lateral tibial and patellar cartilage surfaces for all exercise recovery scans. Each colour represents an individual participant with the black curve representing the mean average trend (loess) of all participants with shaded 95% confidence intervals. Between the baseline scan (timepoint 0) and the first post-exercise scan (timepoint 1), the participant performed a stepping activity dynamically loading the imaged knee for 5 minutes. The first post-exercise $T_{1\rho}$- and T_2-mapping sequences were acquired approximately five and ten minutes after patient positioning, respectively. The last post-exercise $T_{1\rho}$- and T_2-mapping sequences (timepoint 4) were acquired approximately 35 and 40 minutes after patient positioning, respectively. The acquisition of the post-exercise imaging protocol took approximately 45 minutes.

decreases) in $T_{1\rho}$ ($T_{1\rho}$-%SC) and T_2 (T_2-%SC) relaxation time measurements above the determined measurement errors.

Table 3.5: The total number of canonical surface vertices from all four cartilage surfaces and the percentage of surface covered by cartilage regions experiencing changes in $T_{1\rho}$ ($T_{1\rho}$-%SC) and T_22 (T_2-%SC) above the measurement error in response to exercise.

Cartilage Surface	Total Number of Surface Vertices	$T_{1\rho}$-%SC	T_2-%SC
Femoral	3694	8.1	23.0
Lateral Tibial	916	11.4	76.7
Medial Tibial	999	44.0	3.0
Patellar	1093	39.5	36.2

Average %$T_{1\rho}$ change of -7.9 ± 5.5 % and %T_2 change of +2.8 ± 8.6 % were determined from all canonical patellar cartilage areas experiencing a significant change in relaxation times immediately following exercise. For the canonical femoral cartilage surface, average %$T_{1\rho}$ and

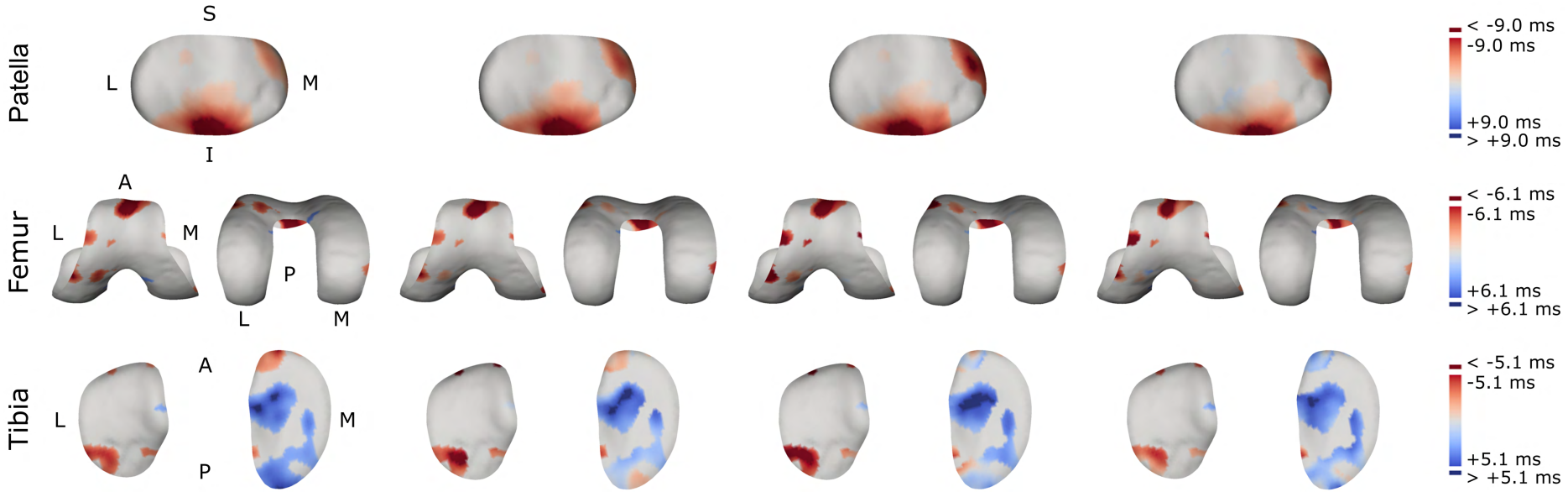

Figure 3.5: Participant-averaged $T_{1\rho}$ difference maps from (top) patellar, (middle) femoral, (bottom) lateral and medial tibial cartilage surfaces represented on canonical model surfaces. The difference maps were calculated by subtracting the average pre-exercise measurement from all four post-exercise recovery measurements (left to right: 1. Post – Pre; 2. Post – Pre; 3. Post – Pre; 4. Post – Pre). Cartilage regions experiencing decreases in $T_{1\rho}$ are specified in red, and regions with an increase in $T_{1\rho}$ compared to the pre-exercise measurement are specified in blue. Only regions experiencing changes larger than the determined thresholds from the repeatability scans are colour-coded. Other areas have been thresholded to zero.

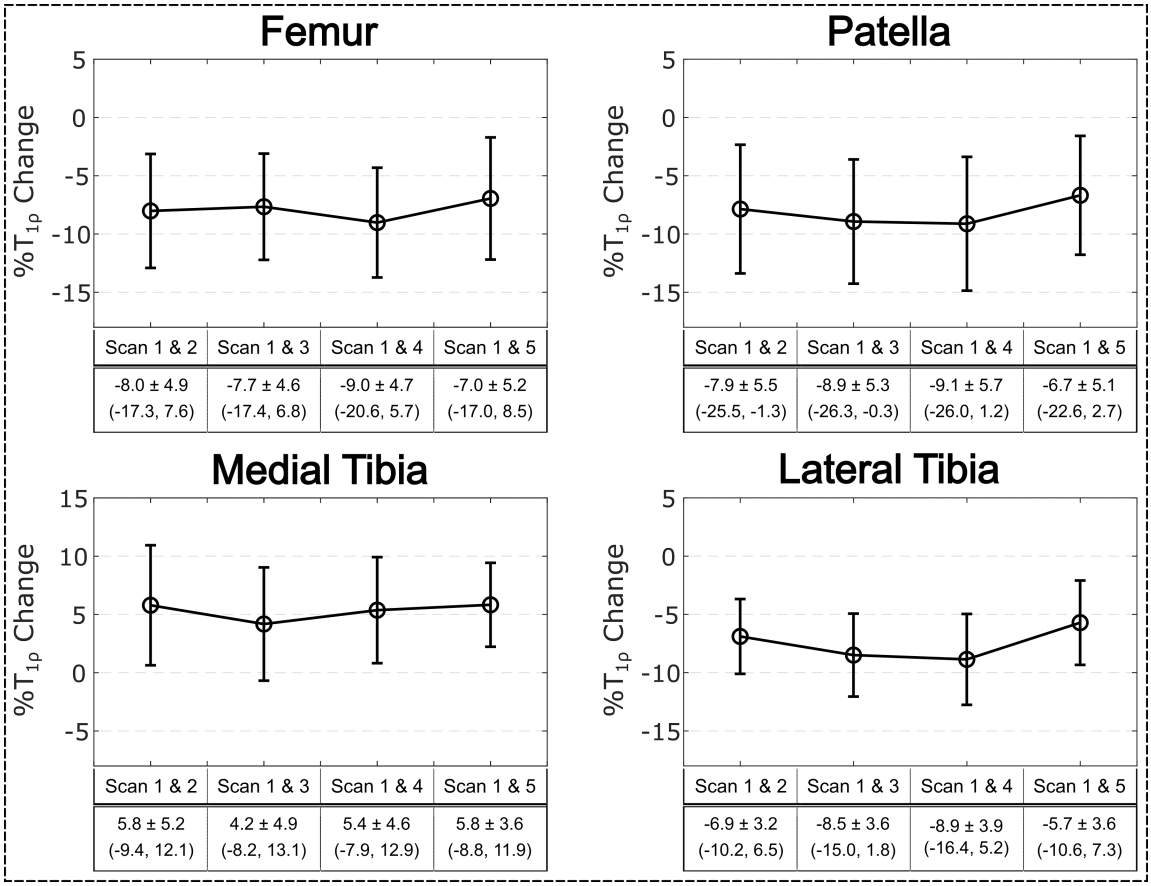

Figure 3.6: Plots showing the normalized change in participant-average femoral (top left), patellar (top right), medial tibial (bottom left) and lateral tibial (bottom right) $T_{1\rho}$ ($\%T_{1\rho}$ change) determined from the four post-exercise measurements (scans 2 – 5) and the one pre-exercise baseline measurement (scan 1). $\%T_{1\rho}$ change at each vertex was calculated according to Equation 3.2.4 and then averaged. The black solid lines represents the collective $\%T_{1\rho}$ change from all areas experiencing a significant change (increase and decrease) between a post-exercise time-point and pre-exercise measurement. Below the plot is a table containing $\%T_{1\rho}$ change mean $\pm$ SD (range) [%] from all vertex-wise calculated normalized changes in the areas experiencing significant variations.

$\%T_2$ changes of -8.0 ± 4.9 % and -5.3 ± 2.3 % were observed in response to exercise, respectively. Average $\%T_{1\rho}$ and $\%T_2$ changes determined from all canonical lateral tibial cartilage regions displaying significant responses to exercise were -6.9 ± 3.2 % and -5.9 ± 2.8 %, respectively. Average medial tibial cartilage $\%T_{1\rho}$ change of +5.8 ± 5.2 % and $\%T_2$ change of +2.8 ± 9.5 % were determined.

The highest negative normalised change of -25.5% was observed in the patellar cartilage $T_{1\rho}$ followed by -17.3% in femoral cartilage $T_{1\rho}$ and -15.0% in lateral tibial cartilage T_2. The largest positive normalised change of +28.4% was displayed in the patellar cartilage T_2 followed by +15.7% in medial tibial cartilage T_2 and +12.1% in medial tibial cartilage $T_{1\rho}$.

When looking at cartilage compositional recovery following exercise and comparing the surface $\%T_{1\rho}$ and $\%T_2$ changes calculated from first post exercise measurements with the $\%T_{1\rho}$ and $\%T_2$ changes determined from last post exercise measurements, patella cartilage $\%T_{1\rho}$ change recovered by 15% while the T_2 'recovered' by 171%. The overall femoral cartilage $\%T_{1\rho}$ change dropped by 13% and the $\%T_2$ change increased by 2% compared to the initial, first post exercise percentage change. While the lateral tibial cartilage $\%T_{1\rho}$ change decreased by 15% of its initial value, the medial tibial $\%T_{1\rho}$ change increased by 1%. The overall $\%T_2$ change of both lateral and medial tibial cartilage increased by 12% and 50% compared to their

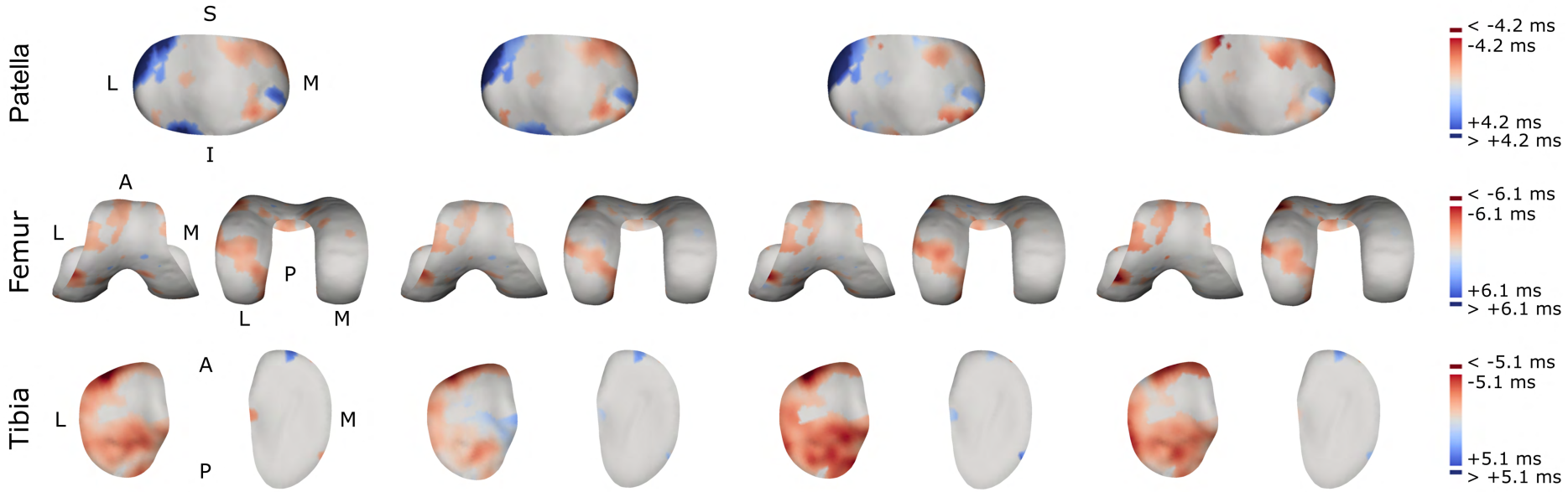

Figure 3.7: Participant-averaged T_2 difference maps from (top) patellar, (middle) femoral, (bottom) lateral and medial tibial cartilage surfaces represented on canonical model surfaces. The difference maps were calculated by subtracting the average pre-exercise measurement from all four post-exercise recovery measurements (left to right: 1. Post – Pre; 2. Post – Pre; 3. Post – Pre; 4. Post – Pre). Cartilage regions experiencing decreases in T_2 are specified in red, and regions with an increase in T_2 compared to the pre-exercise measurement are specified in blue. Only regions experiencing changes larger than the determined thresholds from the repeatability scans are colour-coded. Other areas have been thresholded to zero.

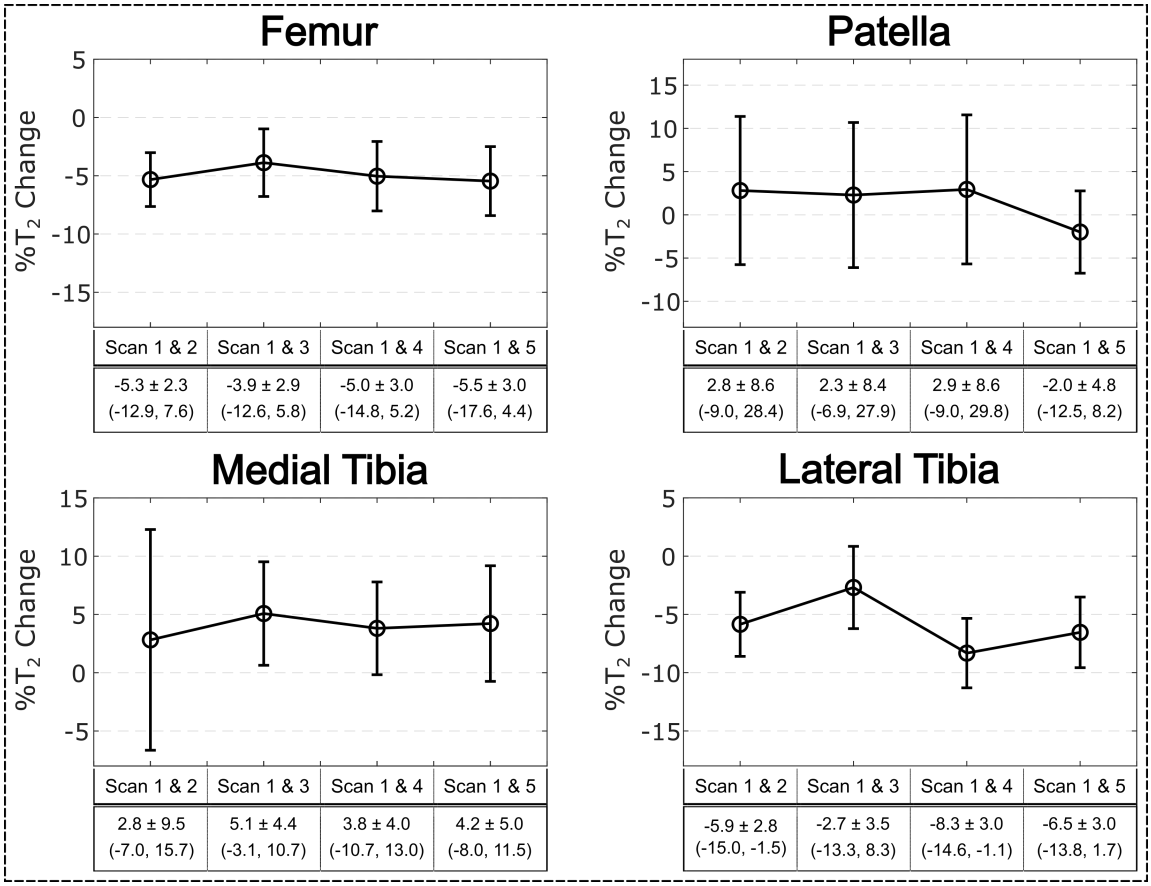

Figure 3.8: Plots showing the normalized change in participant-average femoral (top left), patellar (top right), medial tibial (bottom left) and lateral tibial (bottom right) T_2 ($\%T_2$ change) determined from the four post-exercise measurements (scans 2–5) and the one pre-exercise baseline measurement (scan 1). $\%T_2$ change at each vertex was calculated according to Equation 3.2.4 and then averaged. The black solid lines represents the collective $\%T_2$ change from all areas experiencing a significant change (increase and decrease) between a post-exercise time-point and pre-exercise measurement. Below the plot is a table containing $\%T_2$ change mean $\pm$ SD (range) [%] from all vertex-wise calculated normalized changes in the areas experiencing significant variations.

initial values, respectively.

3.4 Discussion

This work determined the effects of a mild dynamic stepping exercise on the MR relaxation times of cartilage surfaces related to variation in biochemical composition.

The intra-sessional repeatability coefficients-of-variation for $T_{1\rho}$ and T_2 in this study were lower than or comparable to those determined in previous studies [41]. When looking at the surface-averaged $T_{1\rho}$ and T_2 repeatability measurements without knee repositioning, both $T_{1\rho}$ and T_2 were very repeatable on all surfaces. Repositioning of the knee had the greatest effect on the $T_{1\rho}$ relaxation time measurements of patellar cartilage. During repositioning the knee joint experienced bending which could lead to larger changes in cartilage composition at the patellofemoral cartilage contact areas though friction than at the tibiofemoral areas. Averaging of relaxation times over large surfaces could mask these effects on the femoral cartilage surface due to its greater size in comparison to the smaller patellar surface. However, knee repositioning did not show a similarly strong effect on the patellar T_2 relaxation time measure-

ments. This could be a consequence from the time delay (≈10 minutes) required for patient positioning, localisation and $T_{1\rho}$ data acquisition before the T_2 acquisition started and therefore allowing compositional recovery during this time period.

In this study, 3D surface analysis was performed to help gain a better insight into how different cartilage regions respond to and recover from exercise. When averaging the $T_{1\rho}$ and T_2 measurements over the entire femoral, lateral tibial, medial tibial and patellar cartilage surfaces, no statistically significant exercise-related changes were determined when comparing the pre-exercise scan with the first post-exercise scan. As a previous study has also reported, determining mean relaxation time changes from individual slices or across large regions-of-interest may mask significant focal changes [47]. When the individual vertex-wise relaxation times measurements in this study were re-gridded onto a canonical surface, significant exercise-related focal changes in $T_{1\rho}$ and T_2 were observed. Although individual participants showed different cartilage compositional response to the exercise performed, cartilage regions experiencing compositional responses consistent across all participants became evident. By thresholding the exercise-related changes in MR relaxation time measurements with the predetermined threshold limits from the repeatability measurements, cartilage regions undergoing significant responses to the mild dynamic joint-loading activity were highlighted.

Since greater overall normalised changes were seen with $T_{1\rho}$ than with T_2 relaxation time measurements, $T_{1\rho}$ may be a more sensitive biomarker for detecting compositional cartilage responses to joint-loading activities. The $\%T_{1\rho}$ changes of patellar (-7.9%), femoral (-8.0%) and lateral tibial (-6.9%) cartilage and the $\%T_2$ changes of femoral (-5.3%) and lateral tibial (-5.9%) cartilage observed in this study are comparable with those seen in previous studies. Mosher *et al* showed a $\%T_2$ change of approximately -2.5% to -3.2% in femoral and -1.3% to -3.6% in lateral tibial cartilage following a 30-minute running activity [61]. Similarly, Subburaj *et al* demonstrated a $\%T_{1\rho}$ change of -4.1% to -14.3% and a $\%T_2$ change of -3.0% to -9.3% in femoral, tibial and patellar cartilage following running for 30 minutes [107]. The joint movements during the stepping activity performed in this study are comparable to the movements during the stair activity carried out in the study by Chen *et al* [71]. Similarly, the 5-minute stepping activity performed in this study showed a greater effect on patellofemoral cartilage $T_{1\rho}$ relaxation times than on those of femorotibial cartilage, especially in the region of patellofemoral cartilage contact.

We not only observed regions experiencing significant decreases but also significant increases in relaxation time measurements immediately following exercise, especially in medial tibial $T_{1\rho}$ and T_2, and patellar T_2. Farrokhi *et al* also demonstrated a slightly increased $\%T_2$ relaxation time change of 0.3% of healthy patellar cartilage following 50 deep knee bends [62]. Gatti *et al* showed an increased medial femoral $\%T_2$ change after participants bicycled for approximately 45 minutes [110]. Areas of increased normalised change could result from water redistribution rather than expulsion, increasing the water content and decreasing collagen and proteoglycan concentrations in these regions.

Various compositional 'recovery' time-courses were determined for the four different cartilage surfaces. While patellar cartilage volume has been shown to recover in an almost linear fashion following 100 knee bends, we did not observe this linear recovery pattern in patellar cartilage composition [54]. Overall, we only observed a drop in compositional normalised change in four instances ($\%T_{1\rho}$ change of patellar, femoral and lateral tibial cartilage; $\%T_2$ change of patellar cartilage) while in the other four instances ($\%T_{1\rho}$ change of medial tibial cartilage; $\%T_2$ change of femoral, medial and lateral tibial cartilage) an increase in normalised change was observed during the recovery period (post-exercise scan 2 → scan 5). Cartilage

morphology (thickness, volume), independent of cartilage health state, has been shown to recover almost fully in about 45-90 minutes following 30 [65] and 100 knee bends [54] and a 30 minute [112] and 20 km run [60]. Based on our results, the focal compositional changes appear to require more time to return to baseline. More cartilage surfaces experienced some degree of compositional recovery in $T_{1\rho}$ compared to T_2, suggesting that the proteoglycan concentration is recovering faster due to water uptake than the changes in the collagen network after cessation of dynamic joint-loading.

The stepping exercise performed in this study is mild and of short duration. This exercise type was chosen as it is thought to be feasible and extendable for use in patients with early stage knee joint disease and minimal accompanying pain. Knowledge of the effects that deformational loads have on cartilage structure and biochemical composition are important when evaluating clinical imaging studies aiming at determining differences in healthy and diseased cartilage. Differences in cartilage compositional MR relaxation time measurements between healthy and osteoarthritic cartilage have been shown to be in the range of 2% - 13% for $T_{1\rho}$ and 1% - 12% for T_2 for large regional analysis [82, 114, 125]. As the disease-induced compositional changes in cartilage reflected in $T_{1\rho}$ and T_2 measurements can be of the same order, and appear in similar cartilage regions, as exercise-induced changes, it is important to mitigate these effects when conducting clinical OA trials. A 3D surface analysis provides the possibility of spatially localising the deformational and compositional effects of joint loading on articular cartilage and could also assist in determining the regions most prone to exhibit cartilage degeneration [3].

3.4.1 Limitations

As the number of participants in the repeatability and exercise-recovery groups was limited, a larger sample size would increase the precision of the study results. A major limitation to in vivo studies assessing cartilage response to different joint-loading activities is that the compositional behaviour of cartilage cannot be determined immediately after cessation of the exercise but only some short time after as time is required to position the participant back in the MRI system and for acquiring the data. Additionally, the $T_{1\rho}$ and T_2 relaxation time mapping data were not acquired simultaneously but sequentially. Although both sequences are fast spin-echo based sequences, the T_2 mapping was always performed about six minutes after $T_{1\rho}$ during which time further compositional recovery could take place preventing an exact comparison between $T_{1\rho}$ and T_2 results. A sequence capable of simultaneous $T_{1\rho}$ and T_2 acquisition, such as the sequence proposed by Li *et al* [48], could help address this issue.

3.5 Conclusion

We have shown that exercise-related changes in cartilage $T_{1\rho}$ and T_2 relaxation times exceed measurement error and can reliably be determined when using the described 3D-CaSM analysis approach. Based on the results presented here, we hypothesise that mapping of cartilage $T_{1\rho}$ and T_2 relaxation times are measuring dissimilar compositional features as similar cartilage regions showed different $T_{1\rho}$ and T_2 responses to exercise. This lack of correlation between the exercise-induced responses in $T_{1\rho}$ and T_2 warrants further investigation. While complete morphological recovery has previously been shown, the question of when, whether and how the different cartilage regions recover completely from compositional variations following joint

loading activities persists.

Chapter 4

The Optimisation of Deep Neural Networks for Segmenting Multiple Knee Joint Tissues from MRIs

Published as *Research Article* in *Computerized Medical Imaging and Graphics*, 86; December 2020.

4.1 Introduction

Osteoarthritis (OA) is a degenerative disease involving the entire synovial joint [9, 14, 126]. Important risk factors for the development of OA include age, muscle weakness, abnormal joint loading due to joint malalignment or overloading (obesity, high impact sport), and injury to the menisci and ligaments [14, 15, 22]. Distinctive hallmarks of OA include the progressive destruction of articular cartilage structure and alterations in the surrounding joint tissues, including bone, meniscus, ligament and peri-articular muscle. Magnetic Resonance Imaging (MRI) is a commonly used tool to evaluate clinical abnormalities of the knee [127]. Morphological changes due to OA are well demonstrated with MRI [128–132]. Tissue specific masks of the knee joint can be useful for the analysis of OA, especially as automated tools continue to be developed and validated [70, 133–143].

For both clinical and research usage, a significant amount of time is spent manually segmenting images to designate tissue-specific regional masks, also known as regions-of-interest (ROIs). Image masking remains a very significant challenge within medical imaging due to heterogeneity in organ appearance and disease progression and presentation. The segmentation of neighbouring soft tissues such as the cruciate ligaments, cartilages and muscles in the knee joint which have similar image intensities (and therefore poor contrast resolution) is an especially demanding task. ROIs can be generated through manual or semi-manual delineation by a trained reader, or they may be generated automatically using signal thresholding [141], shape [133, 138], atlas [135, 139], or derived from region based [136, 137, 140] approaches, as well as with machine learning approaches [70, 134, 142, 143]. Machine learning methods include unsupervised learning, such as k-means clustering, which segments based on spatial clusters of similar signal intensities in an image [136, 137, 140], or supervised learning by training the algorithm on image masks that have been obtained from any previous masking technique [70, 134, 142, 143]. The number of high-quality label maps for supervised learning is typically

very small, and the performance of a machine learning network trained on a low number of data is limited due to the lack of heterogeneity of images presented during training. Transfer learning may be used to mitigate this by pretraining a network on a large dataset with different but related similarities to the actual task, followed by network refinement on the small dataset [144].

Convolutional neural networks (CNNs), in particular U-Nets [93], have demonstrated their capability to automate the segmentation of musculoskeletal MRIs [70, 145]. Nevertheless, a drawback of this approach with CNNs is that they usually use pixel-wise measures such as the absolute (L1) or square (L2) error loss which can be non-optimal for image data, and, in the case of L2, result in blurry boundaries [146]. In contrast, generative adversarial networks (GANs) [96] learn a similarity measure (feature-wise metric) that adapts to the training task by implementing two competing, or adversarial, neural networks. During adversarial training, one network focusses on image discrimination and guides a second network which focusses on image generation to create "real" images that have a data distribution indistinguishable from the training data distribution. The generator and discriminator are trained simultaneously and competitively in a mini-max game while convergence is achieved when the Nash equilibrium is reached, i.e. no network can improve through further training if one remains unchanged [147].

Conditional GANs (cGANs) modify the GAN approach to learn image-to-image mappings [96, 97]. In comparison to traditional GANs that learn a mapping from random noise to a generated output, cGANs learn a mapping from an observed variable, for example an image to generate an output, such as a label map [96, 97]. cGANs have been used to produce image labels for neurological [148], cardiac [149], abdominal [150], respiratory [151] and musculoskeletal imaging [94, 95]. Liu *et al* used unpaired image-to-image translation with a method called cycle-consistent generative adversarial network (CycleGAN) to perform semantic image segmentation of femorotibial cartilage and bone of the knee joint of unlabelled MRI datasets [95]. The "pix2pix" framework is one cGAN approach that has demonstrated segmentation capability [97]. Semantic segmentation with cGANs, particularly those combining U-Net generators and Markov Random Field discriminators (patch-based discriminators), is relatively unexplored. The method has previously been performed for semantic segmentation of the brain [148]. In Gaj *et al*, a cGAN was used for semantic segmentation of knee cartilage and meniscus but with an image-wise discriminator rather than a patch-wise discriminator [94].

The aim of this study was to implement and evaluate a cGAN for automated semantic segmentation of multiple joint tissues from MR images: the femoral, tibial and patellar bones and cartilage surfaces; the cruciate ligaments; and two selective muscles, the medial vastus and gastrocnemius. Our essential contributions are summarised as followed:

1. Implementation of a cGAN based on the "pix2pix" framework introduced by Isola *et al* using a U-Net generator and a patch-based discriminator for automatic segmentation of multiple knee joint tissues [97]. As far as we know, cGANs have not previously been used for semantic segmentation of the patellar bone and cruciate ligaments, as well as muscles of the knee joint.

2. Evaluating the segmentation performance of the cGAN with different objective functions by combining the cGAN loss with different pixel-wise error losses and modifying the weighting hyperparameter between the cGAN loss and pixel-wise error loss.

3. Assessing the choice of the generator depth and discriminator receptive field size on the performance of the cGAN for multi-tissue segmentation.

4. Quantitative comparison of the cGAN approach with the well-known U-Net approach.

5. Exploring the use of transfer learning for improved segmentation performance of both cGAN and U-Net.

4.2 Methods

4.2.1 Image Datasets

Three image datasets were used for network training and testing; the publicly available SKI10 and OAI ZIB datasets, consisting of 100 and 507 labelled knee MRs, respectively, and a locally acquired dataset of ten segmented knee MRs (Advanced MRI of Osteoarthritis (AMROA) study) [3].

SKI10

The "Segmentation of Knee Images 2010" (SKI10) dataset, consists of approximately 90% 1.5T and 10% 3.0T sagittal MR images using multiple system vendors – GE , Siemens, Philips, Toshiba, and Hitachi [152]. The sequences were varied and included both gradient echo and spoiled gradient echo sequences, commonly with fat suppression. The images were segmented on a slice-by-slice basis by experts from Biomet, Inc., initially through intensity thresholds and thereafter with manual editing. One hundred 3D image datasets of the SKI10 challenge were provided with semi-manual masks of femoral and tibial cartilage and bone. In our study, 70 datasets were used for network training and 30 for network testing.

OAI ZIB

The OAI ZIB dataset is comprised of segmentations of femoral and tibial cartilage and bone of 507 MR imaging volumes from the publicly available Osteoarthritis Initiative dataset ("The Osteoarthritis Initiative," n.d.) [153, 154]. The MR images were acquired on Siemens 3T Trio systems using a 3D double echo steady state (DESS) sequence with water excitation. Outlines of femoral and tibial bone and cartilage were generated using a statistical shape model [138] with manual adjustments performed by experts at Zuse Institute Berlin. The OAI ZIB data covers all degrees of OA (KL 0 – 4), with more cases having severe OA (KL $\geq$ 3) [153]. As with the SKI10 dataset, we split the dataset in 70% (355) for network training and 30% (152) for testing.

AMROA

The locally acquired participant cohort consisted of ten subjects: five healthy volunteers and five patients with mild-to-moderate OA. The patients followed at least one subset of American College of Rheumatology criteria for OA and were recruited between April 2017 to April 2018 (Table 4.1). The healthy volunteers were approximately matched to OA patients for age, sex, and body mass. Network training was performed on data from four subjects with OA and four healthy subjects. Two individuals (one with OA and one healthy) were used as a unique set for test measurements. The number of test individuals was chosen such that roughly 80% of

the data could be used for training. Ethical approval was obtained from the National Research Ethics Service, and all subjects provided written informed consent before participation.

Table 4.1: Participant characteristics showing the mean age, number of males/females (M/F), average body-mass-index (BMI), Kellgren-Lawrence (KL) osteoarthritis score and the number of training/testing set images of the locally acquired dataset. Additionally, the number of participants (N) and training/testing set images of the SKI10 and OAI ZIB datasets are given.

Dataset	Variable	Training Set	Testing Set
Local	N	8	2
	Images	806	171
	Mean Age (years)	53	52
	Sex (M/F)	5/3	0/2
	Mean BMI (kg/m2)	27.8	27.7
	KL (0/2/3)	4/1/3	1/1/0
SKI10	N	70	30
	Images	6133	2626
OAI ZIB	N	355	152
	Images	43814	18517

The source images (Fig. 2A) for each subject were 3D fat-saturated spoiled gradient recalled-echo (3D-FS SPGR) images and were acquired on a 3.0T MRI system (MR750, GE Healthcare, Waukesha, WI, USA) using an 8-channel transmit/receive knee coil (InVivo, Gainesville, FL, USA). The 3D-FS SPGR sequence parameters were: field-of-view=150 x 128 x 136 mm^3, matrix size=512 x 380 x 136 zero-fill interpolated to 512 x 512 x 136, voxel size=0.29 x 0.29 x 1.0 mm^3, TR = 12.5 ms, TE = 2.4 ms, flip angle = 25°, coil acceleration factor (ASSET) = 2, partial Fourier phase encoding = 0.5 (half-NEX), bandwidth = ±11.9 kHz, with fat-suppression.

Semi-manual segmented masks (Fig. 2A) of the patella, tibia, and femur bones as well as of their respective surrounding patellar, tibial and femoral cartilages (Fig. 2b) were created from the 3D-FS SPGR images by a musculoskeletal radiologist with 8 years' experience, using the Stradwin software v5.4a (University of Cambridge Department of Engineering, Cambridge, UK, now freely available as 'StradView' at http://mi.eng.cam.ac.uk/Main/StradView/) [3]. Additionally, masks of the vastus medialis and medial head of gastrocnemius muscles were created. This semi-manual segmentation pipeline consists of sparse manual contour generation (every 2nd-5th sagittal image/2-5 mm) followed by automatic surface triangulation using the regularised marching tetrahedra method. Volume preserving surface smoothing allows creation of an accurate segmentation from relatively sparse manual contours [120]. Manual segmentations of the anterior cruciate ligament (ACL) and posterior cruciate ligament (PCL) were created on the 3D-FS SPGR images using ITK SNAP [155] by a radiologist with 3 years' experience.

4.2.2 Training Data and Masking

Each of the major structures were given a separate image value, i.e., colour, in the segmentation mask, such that the network determined the unique weights to generate a similar regional colour-value from an MR image. On a 256-bit colour-scale, the three bones were stored in the blue colour channel where the femur colour code was 50, tibia was 100, and patella was 150. The cartilages were stored in the green colour channel where the femoral cartilage colour code was 50, the tibial was 100 and the patellar was 150. Additionally, for the AMROA dataset, the muscles were stored in the red colour channel with the medial vastus muscle code set to 100 and the medial gastrocnemius muscle colour code set to 200. The ACL mask was stored in the blue colour channel and the PCL in the green colour channel with both colour codes set to 200.

The MRIs and image masks were converted from the DICOM and NIFTI formats [156], respectively, to a common image format (Portable Network Graphics, PNG) before training. Noise-only images were not used for training or testing, as training a network to fit against zero-valued masks results in a poor constraint. After network training, a tissue- / region-specific Boolean mask was created on the predicted test images by removing prediction values outside of ±20 colour scale units of the tissue specific value. 3D mask predictions were obtained by iterating over the 2D segmented slices.

4.2.3 Network Specifications

This work uses the "pix2pix" framework of a conditional GAN (cGAN) described by Nvidia [97]. The cGAN consists of two deep neural networks, a generator (G) and a discriminator (D). For our task, G learns to translate sagittal MR images of the knee joint (source images x) to semantic segmentation maps ($G(x)$), while D aims to differentiate between the real segmentation map (y) and the synthetically generated.

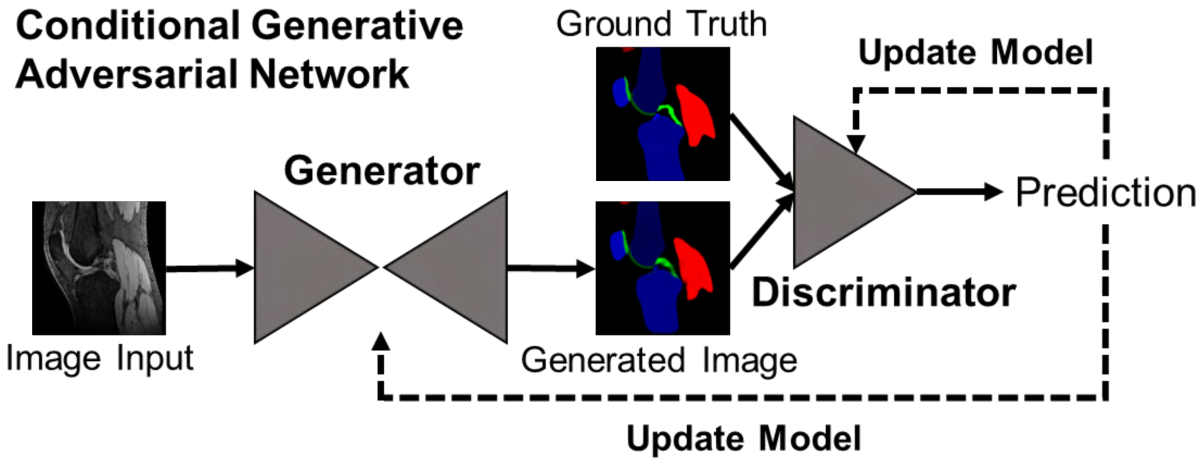

Figure 4.1: Conditional GAN structure. The generator is a U-Net that progressively down-samples / encodes and then up-samples / decodes an input by a series of convolutional layers, with additional skip-connections between each major layer. The generated, 'fake' segmentation image is then fed together with the ground truth segmentation image into a discriminator network (PatchGAN (Isola et al., 2017)) that gives its prediction of whether the generated image is a 'real' representation of the ground truth image, or not. A detailed description of the network architecture can be found in the Appendix.

The structure of a cGAN is illustrated in Figure 4.1. The loss function for this cGAN is

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}\left[\log D(x, y)\right] + \mathbb{E}_{x}\left[\log(1 - D(x, G(x)))\right] \tag{4.2.1}$$

The loss function describes how G is minimized against a maximised D. Since both optimisation processes are dependent on each other, convergence is achieved by reaching a saddle point (simultaneously minimum / maximum for both networks' cost) rather than a minimum. The loss also incorporates a L1 distance to reduce image blurring and ensure that the generated image from $G(x)$ are not significantly different from the target image y [97, 157]. This L1 loss is given by

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}\left[\|y - G(x)\|_1\right] \tag{4.2.2}$$

The overall objective of the cGAN is to find the optimal solution to

$$G* = \arg\min_{G}\max_{D} \mathcal{L}_{cGAN}(G, D) + \lambda\mathcal{L}_{L1}(G) \tag{4.2.3}$$

with λ being a hyper-parameter used for balancing the two losses (Regmi and Borji, 2018).

The cGAN used in this work utilises the U-Net encoder-decoder architecture for the generator, which is frequently used for image segmentation problems [93]. The generator was trained to generate images that are indistinguishable from a target image (i.e., the segmented map). Spatial consistency of the data is not guaranteed with a U-Net segmented map, which can cause inaccurate boundaries [93]. However, adversarial losses in the discriminator regulate and therefore increase the accuracy to higher order shapes [158].

We modified the U-Net generator from the "pix2pix" network by increasing the input layer to be able to train on 512 x 512 resolution images. For this an additional Convolution-BatchNorm-leakyReLU layer was inserted in the encoding and a Convolution-BatchNorm-ReLU layer in the decoding network part.

The discriminator is a patch-based fully convolutional neural network, PatchGAN [159, 160], which models the image as a Markov random field. It performs a convolutional patch-wise (N x N) classification with all the outputs in the patch averaged and taken as the output of D. D is therefore less dependent on distant pixels/voxels beyond a "patch diameter" and is a form of neighbouring texture loss. The PatchGAN can be applied to arbitrarily large images, due to a fixed size of the patch.

To analyse the cGANs performance we compared it to the performance of a U-Net network, which is widely used for image segmentation processes. We used the cGAN generator network as the U-Net network to maintain an effective comparison.

The networks were implemented using PyTorch (Torch v1.0.1) and all training was performed on a Nvidia P6000 GPU card (3840 CUDA cores, 24 GB GDDR5X). The training phase of optimisation was performed as described by the "pix2pix" network, using stochastic gradient descent to minimise D(x,y) and stochastic gradient ascent to maximise D(x,G(x)). The Adam solver was used with a learning rate 0.0002 and momentum parameters, β_1 = 0.5, β_2 = 0.999. We introduced random noise (jitter) during training by resizing the input images to 542 x 542 using bi-cubic interpolation followed by random cropping back to 512 x 512.

A detailed description of the network architectures can be found in the Appendix.

4.2.4 Segmentation Evaluation Metrics

The Sørensen–Dice Similarity Coefficient (DSC) [89, 161] was used to evaluate the overlap between the generated segmentation and the manual segmentation. The DSC ranges between 0 and 1, with 0 representing no overlap and 1 complete overlap between the two sets. DSC is defined as twice the size of the intersect divided by the sum of the sizes of two sample sets, given as

$$DSC = \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{4.2.4}$$

for Boolean metrics. For the experiments involving the SKI10 and OAI ZIB datasets, the volumetric overlap error (VOE) and the boundary distance-based metric average surfaces distance (ASSD) were determined to assess segmentation accuracy and allow an appropriate comparison with previous studies using these datasets. The VOE can be calculated as

$$VOE = 1 - \frac{|X \cap Y|}{|X \cup Y|} \tag{4.2.5}$$

with small values for VOE expressing greater accuracy. The ASSD is expressed in mm and is defined as

$$ASSD = \frac{1}{N_X + N_Y}\left(\sum_{i=1}^{N_X} D_X(y) + \sum_{i=1}^{N_Y} D_Y(x)\right) \tag{4.2.6}$$

where $D_X(y) = \min_{x \in X} \|y - x\|$ is the distance of a voxel y to a surface X and $\|\cdot\|$ denotes the Euclidean norm.

4.2.5 Evaluation of Network Characteristics

This section aims at evaluating and adjusting specific network characteristics towards improving overall network performance, for both cGAN and U-Net. All networks in this section were trained for 100 epochs and all cGANs with a 70 x 70 PatchGAN discriminator unless otherwise stated.

Evaluation of Network Objective Function:

We evaluated the cGANs performance with different objective functions by combining the cGAN loss with different pixel-wise error losses. In this work the cGAN is tasked to output a segmentation map of multiple tissues having different features and locations in the input MR image. We assessed the shortcomings and strengths of including the $\mathcal{L}_{L1}$, $\mathcal{L}_{L2}$ and Smooth L1 ($\mathcal{L}_{SmL1}$) [162] loss functions in the cGAN objective. The $\mathcal{L}_{L2}$ loss and $\mathcal{L}_{SmL1}$ loss are given by

$$\mathcal{L}_{L2}(G) = \mathbb{E}_{x,y}\left[\|y - G(x)\|_2^2\right] \tag{4.2.7}$$

$$\mathcal{L}_{SmL1}(G) = \begin{cases} 0.5 \cdot \mathbb{E}_{x,y}\left[\|y - G(x)\|_2^2\right], \text{ if } |y - G(x)| < 1 \\ \mathbb{E}_{x,y}\left[\|y - G(x))\|_1 - 0.5\right], \text{otherwise} \end{cases} \tag{4.2.8}$$

Furthermore, the weighting hyperparameter λ between the cGAN loss and pixel-wise error loss was changed to vary the balance between the two task losses. λ = 0.01, 1, 100 and 10000 were investigated. Network training with the cGAN loss alone (λ = 0) was additionally performed and evaluated.

We also trained the U-Net with the same three different pixel-wise error losses ($\mathcal{L}_{L1}$, $\mathcal{L}_{L2}$ and $\mathcal{L}_{SmL1}$) as the cGAN to maintain an effective comparison.

Evaluation of Altering the Loss Objective during Training:

After obtaining initial results, we observed that the cGAN was unable to segment muscle tissues, independent of the objective function trained on. Therefore, we decided to explore the effect of varying the loss objective during training. For this, we trained a cGAN with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$ loss and a U-Net with $\mathcal{L}_{L2}$ loss for 50 epochs and then changed the loss functions for the ensuing 50 epochs to $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ and $\mathcal{L}_{L1}$, respectively.

Evaluation of the Generator Depth:

We analysed the effect of changing the depth of the generator network on the cGANs and U-Nets quantitative performance. In addition to the generator down-sampling the input through nine convolutional networks, we tested a generator consisting of seven and five convolutions during down-sampling. Furthermore, we assessed the quantitative performance of the generator network with different numbers feature channels. We compared networks starting with different minimum number of feature channels (16, 32, 64 and 128) and thus end at different maximum numbers of feature channels (128, 256, 512 and 1024). All cGANs were trained with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss with λ = 100 and all U-Nets with the $\mathcal{L}_{L1}$ loss. Detailed descriptions of the generator network architectures can be found in the Appendix.

Evaluation of the PatchGAN Receptive Field Size:

We evaluated the effect of changing the PatchGAN receptive field size on the cGANs qualitative (artefact emergence) and quantitative (segmentation accuracy) performance. In addition to the 70 x 70 PatchGAN, we tested a 1 x 1 (PixelGAN), 34 x 34 and 286 x 286 PatchGAN. All cGANs were trained with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss with λ = 100. Detailed descriptions of the discriminator network architectures can be found in the Appendix.

Evaluation of Transfer Learning:

Since the AMROA dataset only comprises of a low number of subjects (N=8) for training, we assess the influence of transfer learning on network performance, by initially training both a cGAN ($\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$) and a U-Net ($\mathcal{L}_{L1}$) for 20 epochs on the larger SKI10 and OAI ZIB training datasets separately followed by network fine-tuning for 80 epochs on the smaller AMROA training set. Additionally, a cGAN and a U-Net were trained for 20 epochs on the AMROA training dataset followed by network refinement training for 80 epochs on either the SKI10 or OAI ZIB training set to analyse the potential segmentation improvement of SKI10 and OAI ZIB. Network performance evaluations were performed using AMROA, SKI10 and OAI ZIB testing datasets. As determined from the previous sections, the cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss objective (λ = 100) and a 1 x 1 PixelGAN as well as the U-Net trained

with the $\mathcal{L}_{L1}$ loss objective achieved the highest segmentation accuracies for most knee joint tissues segmented in the AMROA dataset and were used in this section.

4.3 Results and Discussion

4.3.1 Network Training and Testing

Semi-manual segmentation of the AMROA images by the reader required ≈30 minutes per subject-volume. cGAN training was performed in 80 seconds/epoch for the AMROA training dataset, and 390 seconds/epoch for the SKI10 dataset. U-Net training was performed in 45 seconds/epoch for the AMROA training dataset, and 185 seconds/epoch for the SKI10 dataset. Segmentation post-training on a single slice was processed in ≈0.13s. The highlights of the upcoming sections are:

- Section 4.3.2: The U-Net trained with $\mathcal{L}_{L1}$ loss objective outperformed the cGANs and the U-Nets trained with different loss objectives in the segmentation performance of most knee joint tissues.
- Section 4.3.3: Altering the network objective function midway through cGAN and U-Net training lead to unanticipated but advantageous results. This variation resulted in improved segmentation performances of several tissues and the cGANs capability to segment muscle tissue, which previously had not been possible with non-altered objective function training.
- Section 4.3.4: The cGAN and U-Net trained with nine convolutions/transpose convolutions in the networks encoding/decoding parts and a minimum feature channel change of 64 achieved the highest segmentation accuracies for most knee joint tissues annotated.
- Section 4.3.5: The greatest improvements in segmentation performance of the cGAN was achieved by reducing the receptive field size of the discriminator network. This resulted in segmentation accuracies equivalent to those of the U-Net.
- Section 4.3.6: Transfer learning not only increased segmentation accuracy of some tissues of the fine-tuned dataset, but also increased the network's capacity to maintain segmentation capabilities for the pretrained dataset.
- Section 4.3.7: Overall, the cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss objective (λ = 100) and a 1 x 1 PixelGAN as well as the U-Net trained with the $\mathcal{L}_{L1}$ loss objective achieved comparable and the highest segmentation accuracies for most knee joint tissues segmented.

4.3.2 Evaluation of Network Objective Function

The quantitative results of assessing the impact of combining the cGAN objective with three different pixel error losses with varying weightings λ on the cGANs segmentation performance are in Table 4.2, with the qualitative results depicted in Figure 4.2B. The cGANs trained with larger values for λ (λ = 100 and 10000) achieved the highest segmentation performance for all tissues and the produced segmentation maps were less affected by artefacts compared to the

cGANs trained with λ = 0.01 and 1. For instance, the images from the networks trained with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ (λ = 0.01), $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$ (λ = 1) and $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{SmL1}$ (λ = 1) had artefacts where the networks seem to detect bone or cartilage structures where there were none in the original MR input image. By increasing the weighting hyperparameter λ, more emphasis is put on the pixel error losses to guide the network to produce more accurate representations of the ground truth segmentation map and reduces these artefacts. However, the influence of GAN loss diminishes with very large values for λ with the discriminator having minimal effect on generator training.

The qualitative results of training a U-Net with different pixel error losses are presented in Figure 4.2C while the quantitative results are listed in Table 4.3. The U-Net trained with $\mathcal{L}_{L1}$ loss objective achieves the highest accuracy for all tissues compared to $\mathcal{L}_{L2}$ and $\mathcal{L}_{SmL1}$ loss except for the muscle tissues. Muscle tissues appeared on the majority of 2D MR knee images seen by the network during training, however we only segmented two selective medial muscles in the AMROA dataset due to time constraints. It is interesting to note that although the U-Net trained with $\mathcal{L}_{L1}$ was not able to capture the medial head of gastrocnemius and vastus medialis muscles, the cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ objective (λ = 10000) was. Simple absolute difference ($\mathcal{L}_{L1}$) was not capable of differentiating lateral muscle textures from medial. The U-Nets trained with $\mathcal{L}_{L2}$ and $\mathcal{L}_{SmL1}$ losses were capable of segmenting the selective muscles with high accuracies as they are penalised more by the squaring term in their loss objectives when the difference between ground truth and model predictions are large. Interestingly, although the patella bone and cartilage only appear on very few slices in a 3D dataset, and ACL and PCL on even fewer, the U-Net with $\mathcal{L}_{L1}$ segmented these tissues better than the $\mathcal{L}_{L2}$ and $\mathcal{L}_{SmL1}$ ($\mathcal{L}_{L2}$: $\text{DSC}_{PBone} < 0.2\%$, $\text{DSC}_{PCartilage} < 5.3\%$, $\text{DSC}_{ACL} < 15.2\%$, $\text{DSC}_{PCL} < 21.3\%$; $\mathcal{L}_{SmL1}$: $\text{DSC}_{PBone} < 0.4\%$, $\text{DSC}_{PCartilage} < 6.0\%$, $\text{DSC}_{ACL} < 6.9\%$, $\text{DSC}_{PCL} < 17.8\%$). This could be explained by the cruciate ligament and patellar tissues either being present or not on a 2D training image and the network is not being constrained to only segment medial tissues. Overall, the U-Net with $\mathcal{L}_{L1}$ produced sharper boundaries, especially for the smaller ligament structures, as compared to the segmentation maps produced by U-Nets trained with $\mathcal{L}_{L2}$ and $\mathcal{L}_{SmL1}$, in which the boundaries are more diffused.

We decided to assess the model's performance when including noise-only images in the testing dataset as we excluded them during model training, and this might limit the models' use in a clinical setting. This effect was only evaluated for a the cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ (λ = 100) objective function and the U-Net trained with the $\mathcal{L}_{L1}$ loss objective. The quantitative results are listed in Table 4.4 with qualitative results displayed in Figure 4.3. Both networks showed comparable segmentation performances after testing with noise-only images with percentage differences (%-Diff) of the DSC for all segmented tissues $\leq 2.3\%$. Including noise-only images into the testing set had greater effects on the cGAN DSC of the medial vastus muscle (VM muscle) (%-Diff = 1.5%), the ACL (%-Diff = 1.6%) and the PCL (%-Diff = 1.9%) as well as on the U-Net DSC of the ACL (%-Diff = 2.3%). These higher differences could be explained by the lower segmentation capability of these structures by the cGAN and U-Net models to begin with (cGAN: DSC_{VM} muscle: 0.113 vs 0.098, DSC_{ACL}: 0.577 vs 0.593; DSC_{PCL}: 0.073 vs 0.092; U-Net: DSC_{ACL}: 0.643 vs 0.620). Furthermore, the larger %-Diff in the DSC of the VM muscle is caused by the cGAN model irregularly segmenting VM muscle tissues on noise only images (Figure 4.3B).

Table 4.2: Results of the Network Objective Function: cGAN. The influence of mixing the cGAN objective with different pixel-wise error losses and varying their significance by changing the weighting hyperparameter λ on the segmentation performance of the proposed cGAN was assessed. Highest network scores achieved for each tissue are in bold.
Training and testing were performed on the AMROA training and testing datasets, respectively.
Results are presented as mean ± standard deviation.
Abbreviations: F Bone – femoral bone, T Bone – tibial bone, P Bone – patellar bone, F Cartilage – femoral cartilage, T Cartilage – tibial cartilage, P Cartilage – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient

Network Objective Function Results											
cGAN											
Pixel Loss	λ	**F Bone**	**T Bone**	**P Bone**	**F Cartilage**	**T Cartilage**	**P Cartilage**	**VM Muscle**	**GM Muscle**	**ACL**	**PCL**
		DSC	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**
L1	0	0.931 ± 0.020	0.864 ± 0.008	0.911 ± 0.036	0.774 ± 0.030	0.717 ± 0.108	0.872 ± 0.030	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	0.01	0.900 ± 0.018	0.890 ± 0.031	0.912 ± 0.002	0.727 ± 0.023	0.715 ± 0.060	0.850 ± 0.048	0.000 ± 0.000	0.000 ± 0.000	0.509 ± 0.009	0.171 ± 0.208
	1	0.899 ± 0.014	0.856 ± 0.010	0.807 ± 0.060	0.465 ± 0.037	0.666 ± 0.022	0.426 ± 0.098	0.611 ± 0.181	0.595 ± 0.054	0.000 ± 0.000	0.000 ± 0.000
	100	0.918 ± 0.011	0.948 ± 0.018	0.928 ± 0.002	0.812 ± 0.002	0.748 ± 0.042	**0.863 ± 0.043**	0.113 ± 0.085	0.000 ± 0.000	0.577 ± 0.020	0.073 ± 0.103
	10000	**0.968 ± 0.006**	0.944 ± 0.026	0.917 ± 0.008	**0.875 ± 0.021**	**0.810 ± 0.036**	0.840 ± 0.065	0.879 ± 0.036	0.793 ± 0.080	0.432 ± 0.237	0.338 ± 0.386
L2	0.01	0.902 ± 0.004	0.915 ± 0.003	0.923 ± 0.005	0.750 ± 0.002	0.740 ± 0.079	0.834 ± 0.077	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	1	0.902 ± 0.046	0.902 ± 0.008	0.902 ± 0.044	0.741 ± 0.004	0.736 ± 0.033	0.838 ± 0.041	0.000 ± 0.000	0.000 ± 0.000	0.149 ± 0.104	0.002 ± 0.002
	100	0.928 ± 0.015	0.939 ± 0.007	0.921 ± 0.022	0.768 ± 0.016	0.752 ± 0.049	0.862 ± 0.039	0.001 ± 0.001	0.000 ± 0.000	**0.652 ± 0.094**	0.101 ± 0.074
	10000	0.952 ± 0.000	**0.950 ± 0.015**	0.923 ± 0.001	0.828 ± 0.043	0.684 ± 0.092	0.832 ± 0.054	0.814 ± 0.145	0.856 ± 0.121	0.440 ± 0.084	0.293 ± 0.358
SmL1	0.01	0.914 ± 0.034	0.902 ± 0.003	0.920 ± 0.011	0.726 ± 0.007	0.729 ± 0.042	0.762 ± 0.068	0.000 ± 0.000	0.000 ± 0.000	0.343 ± 0.066	0.000 ± 0.000
	1	0.884 ± 0.044	0.912 ± 0.006	0.926 ± 0.013	0.740 ± 0.014	0.732 ± 0.044	0.829 ± 0.067	0.055 ± 0.007	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	100	0.903 ± 0.019	0.944 ± 0.006	**0.936 ± 0.003**	0.776 ± 0.035	0.741 ± 0.066	0.857 ± 0.029	0.031 ± 0.044	0.070 ± 0.100	0.578 ± 0.053	0.044 ± 0.052
	10000	0.951 ± 0.002	0.946 ± 0.018	0.935 ± 0.015	0.825 ± 0.035	0.738 ± 0.047	0.797 ± 0.088	**0.914 ± 0.001**	**0.837 ± 0.146**	0.261 ± 0.073	**0.374 ± 0.341**

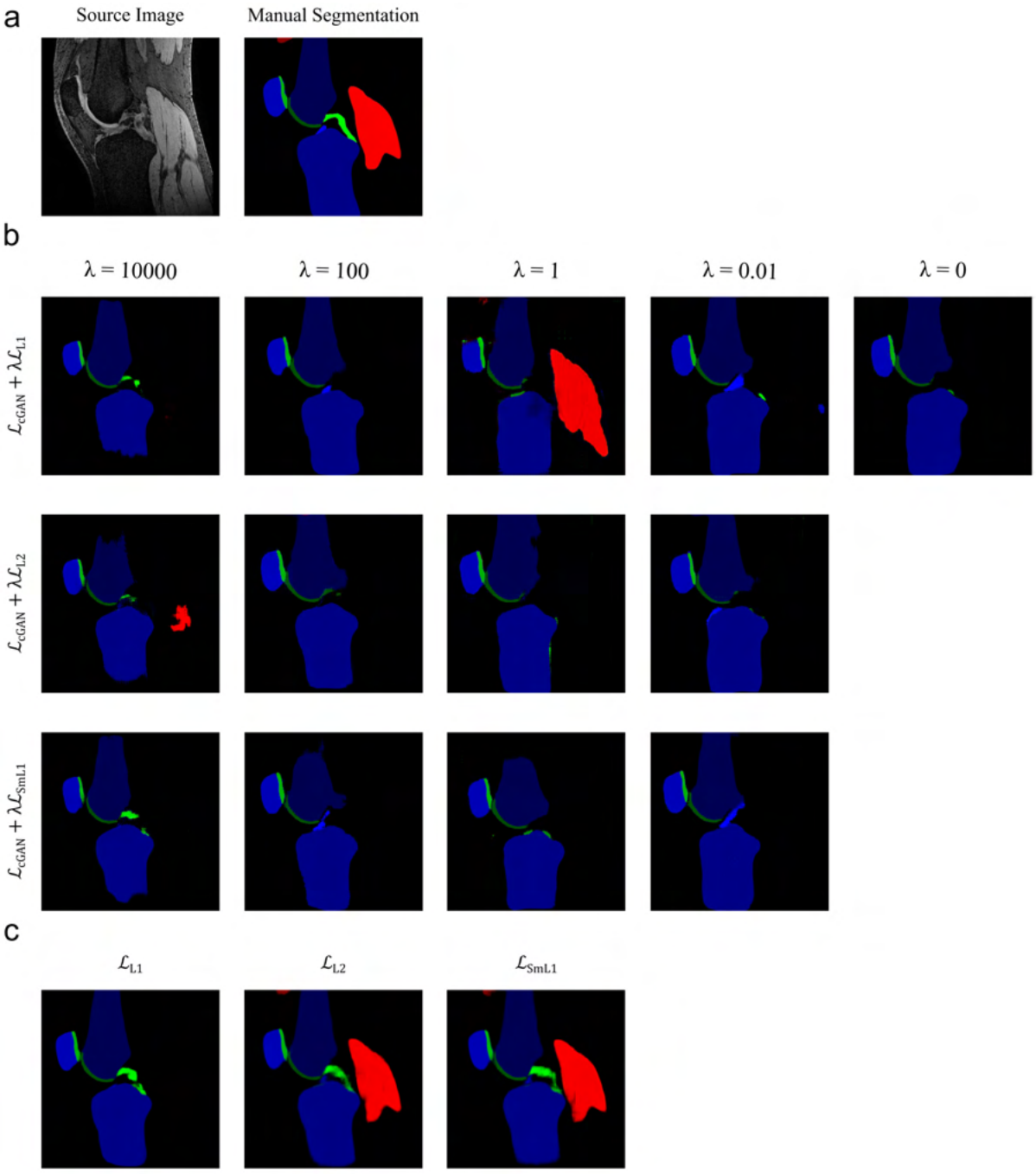

Figure 4.2: Results of Network Objective Function. Qualitative results of B) training a cGAN with different objective functions by combining the cGAN loss with different pixel-wise error losses with varying weightings and C) training a U-Net with different pixel-wise error losses.

Table 4.3: Results of the Network Objective Function: U-Net. The influence of different pixel-wise error losses on the segmentation performance of the U-Net was assessed. Highest network scores achieved for each tissue are in bold.
Training and testing were performed on the AMROA training and testing datasets, respectively.
Results are presented as mean ± standard deviation.
Abbreviations: F Bone – femoral bone, T Bone – tibial bone, P Bone – patellar bone, F Cartilage – femoral cartilage, T Cartilage – tibial cartilage, P Cartilage – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient

Network Objective Function Results										
	U-Net									
Pixel Loss	**F Bone**	**T Bone**	**P Bone**	**F Cartilage**	**T Cartilage**	**P Cartilage**	**VM Muscle**	**GM Muscle**	**ACL**	**PCL**
	DSC	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**
L1	**0.972 ± 0.006**	**0.960 ± 0.001**	**0.941 ± 0.010**	**0.886 ± 0.007**	**0.834 ± 0.010**	**0.890 ± 0.034**	0.000 ± 0.000	0.000 ± 0.000	**0.643 ± 0.153**	**0.641 ± 0.008**
L2	0.950 ± 0.007	0.957 ± 0.009	0.939 ± 0.003	0.831 ± 0.020	0.723 ± 0.068	0.837 ± 0.051	0.888 ± 0.000	0.881 ± 0.021	0.491 ± 0.136	0.428 ± 0.196
SmL1	0.953 ± 0.001	0.953 ± 0.009	0.937 ± 0.004	0.843 ± 0.021	0.771 ± 0.036	0.830 ± 0.088	**0.894 ± 0.002**	**0.910 ± 0.045**	0.574 ± 0.230	0.463 ± 0.174

Table 4.4: Results of additionally testing on noise only images. The influence of including noise only images in the testing set on the overall segmentation performance of a cGAN trained with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ (λ = 100) loss objective and a U-Net trained with $\mathcal{L}_{L1}$ objective. Training was performed on the AMROA training dataset without noise only images.
Abbreviations: F Bone – femoral bone, T Bone – tibial bone, P Bone – patellar bone, F Cartilage – femoral cartilage, T Cartilage – tibial cartilage, P Cartilage – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient, %-Diff – absolute percentage difference

Influence of Noise Only Images										
cGAN										
Testing	**F Bone**	**T Bone**	**P Bone**	**F Cartilage**	**T Cartilage**	**P Cartilage**	**VM Muscle**	**GM Muscle**	**ACL**	**PCL**
	DSC	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**
No Noise	0.918 ± 0.011	0.948 ± 0.018	0.928 ± 0.002	0.812 ± 0.002	0.748 ± 0.042	0.863 ± 0.043	0.113 ± 0.085	0.000 ± 0.000	0.577 ± 0.020	0.073 ± 0.103
With Noise	0.925 ± 0.012	0.946 ± 0.017	0.928 ± 0.004	0.810 ± 0.003	0.752 ± 0.045	0.858 ± 0.054	0.098 ± 0.114	0.000 ± 0.000	0.593 ± 0.028	0.092 ± 0.131
%-Diff	0.7	0.2	0.0	0.2	0.4	0.5	1.5	0.0	1.6	1.9
U-Net										
Testing	**F Bone**	**T Bone**	**P Bone**	**F Cartilage**	**T Cartilage**	**P Cartilage**	**VM Muscle**	**GM Muscle**	**ACL**	**PCL**
	DSC	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**
No Noise	0.972 ± 0.006	0.960 ± 0.001	0.941 ± 0.010	0.886 ± 0.007	0.834 ± 0.010	0.890 ± 0.034	0.000 ± 0.000	0.000 ± 0.000	0.643 ± 0.153	0.641 ± 0.008
With Noise	0.968 ± 0.001	0.957 ± 0.009	0.938 ± 0.016	0.885 ± 0.004	0.833 ± 0.010	0.894 ± 0.026	0.000 ± 0.000	0.000 ± 0.000	0.620 ± 0.156	0.643 ± 0.025
%-Diff	0.4	0.3	0.3	0.1	0.1	0.4	0.0	0.0	2.3	0.2

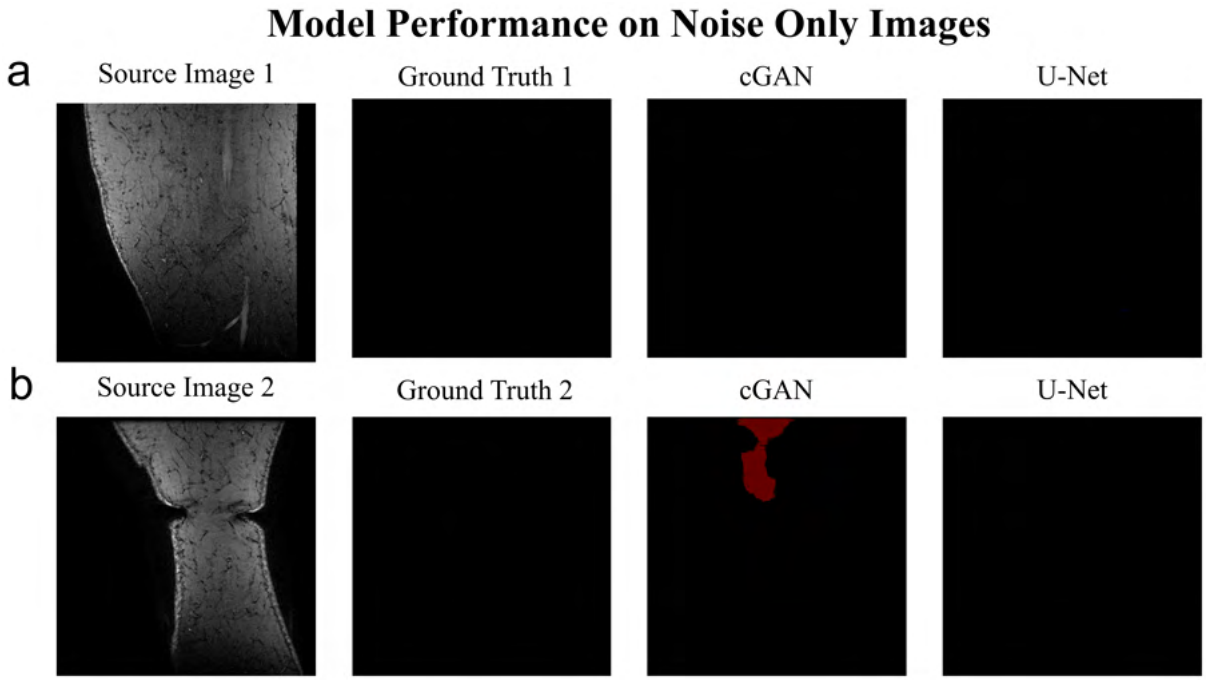

Figure 4.3: Results of testing on noise only images. Assessing the segmentation performance of a cGAN trained with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ (λ = 100) loss objective and a U-Net trained with $\mathcal{L}_{L1}$ objective and tested on noise only images. Training was performed on the AMROA training dataset without noise only images. A) and B) are two example results of testing the models on noise only source images and comparing to ground truth segmentation maps.

4.3.3 Evaluation of Altering Loss Objective during Training

Figure 4.4 compares the qualitative results and Table 4.5 compares the DSCs obtained from a cGAN and a U-Net, in which the objective functions were changed midway through training to the cGANs and U-Nets trained with non-altered objective functions. Training a cGAN with varied loss objective ($\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2} \rightarrow \mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$) notably reduced its ability to segment the ACL, however considerably improved its segmentation performance on the medial vastus and gastrocnemius muscles, as well as PCL, compared to the other cGANs ($\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ and $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$). The images in Figure 4.4B show the improvements in muscle segmentation with the cGAN trained with varied loss objective. This was a surprising result as neither the cGAN trained with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ nor with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$ alone were able to segment muscle. Looking at the different training epochs of the cGAN trained with varied loss, during $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$ no muscle tissue was being semantically segmented. However, when changing to $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ and between training epochs 50 and 60, the network started segmenting muscle tissue (Figure 4.5). After the initial 50 epochs of $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$ training, the cGANs weights must have been favourable for continuing training with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ to additionally semantically segment muscle tissue.

The U-Net trained with altered objective function ($\mathcal{L}_{L2} \rightarrow \mathcal{L}_{L1}$) also showed notable improvements in the segmentation performance of the medial vastus and gastrocnemius muscles while the segmentation scores of the other knee tissues remained comparable with those of the other U-Nets ($\mathcal{L}_{L1}$ and $\mathcal{L}_{L2}$). Figure 4.4C qualitatively compares the results of a U-Net trained with altered loss objective to those of the U-Nets trained with a single, non-altered loss objective. As mentioned in the corresponding method section, this idea came after reviewing a few initial training results. While the U-Net trained with the $\mathcal{L}_{L1}$ objective was not able to segment

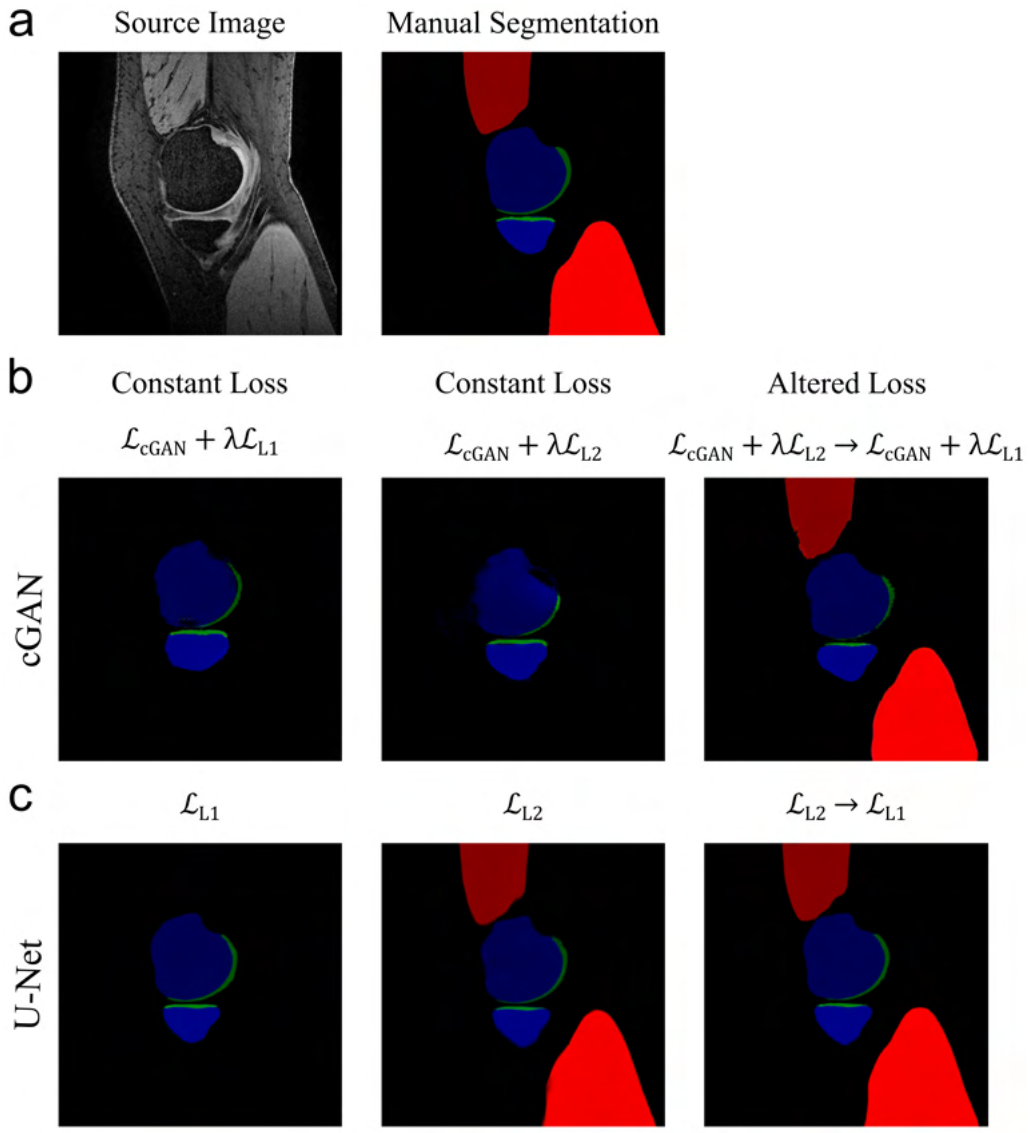

Figure 4.4: Results of Altering the Loss Objective during Training. Assessing the influence of varying the objective function halfway during cGAN and U-Net training on their segmentation performance with comparison to the respective cGANs and U-Nets trained with constant loss function.

the medial vastus and gastrocnemius muscles after training, the U-Net with the $\mathcal{L}_{L2}$ loss objective was. However, these images were slightly blurrier, and the segmentation accuracy of the remaining tissues was poorer than compared to $\mathcal{L}_{L1}$. By varying the loss objective during training, the strengths of $\mathcal{L}_{L2}$ and $\mathcal{L}_{L1}$ were combined. We decided to first train the network with $\mathcal{L}_{L2}$ loss to capture all tissues and then to change to $\mathcal{L}_{L1}$ halfway through training to make the images sharper and increase segmentation accuracy. This method created a more proficient network capable of segmenting all tissues with higher or comparable accuracies to the networks trained with non-altered loss objectives.

Table 4.5: Results of Altering the Loss Objective during Training. Assessing the influence of altering the loss objective function during training on the segmentation performance of the proposed cGAN and U-Net. A cGAN was trained with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$ objective and a U-Net with $\mathcal{L}_{L2}$ objective for 50 epochs followed by a further 50 epochs training with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ and $\mathcal{L}_{L1}$ objectives, respectively. Segmentation performances are compared with the previously trained cGANs ($\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ and $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$; λ = 100; 100 epochs) and U-Nets ($\mathcal{L}_{L1}$ and $\mathcal{L}_{L2}$; 100 epochs). Highest network scores achieved for each tissue are in bold.
Training and testing were performed on the AMROA training and testing datasets, respectively.
Abbreviations: FB – femoral bone, TB – tibial bone, PB – patellar bone, FC – femoral cartilage, TC – tibial cartilage, PC – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient

Altering the Loss Objective during Training Results										
cGAN										
Network Loss Objective	**F Bone DSC**	**T Bone DSC**	**P Bone DSC**	**F Cartilage DSC**	**T Cartilage DSC**	**P Cartilage DSC**	**VM Muscle DSC**	**GM Muscle DSC**	**ACL DSC**	**PCL DSC**
$\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$	0.918 ± 0.011	**0.948 ± 0.018**	**0.928 ± 0.002**	**0.812 ± 0.002**	0.748 ± 0.042	**0.863 ± 0.043**	0.113 ± 0.085	0.000 ± 0.000	0.577 ± 0.020	0.073 ± 0.103
$\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$	0.928 ± 0.015	0.939 ± 0.007	0.921 ± 0.022	0.768 ± 0.016	0.752 ± 0.049	0.862 ± 0.039	0.001 ± 0.001	0.000 ± 0.000	**0.652 ± 0.094**	0.101 ± 0.074
$\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2} \rightarrow \mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$	**0.936 ± 0.007**	0.938 ± 0.021	0.884 ± 0.078	0.800 ± 0.021	**0.760 ± 0.035**	0.855 ± 0.031	**0.739 ± 0.010**	**0.772 ± 0.005**	0.115 ± 0.032	**0.392 ± 0.128**
U-Net										
Network Loss Objective	**F Bone DSC**	**T Bone DSC**	**P Bone DSC**	**F Cartilage DSC**	**T Cartilage DSC**	**P Cartilage DSC**	**VM Muscle DSC**	**GM Muscle DSC**	**ACL DSC**	**PCL DSC**
$\mathcal{L}_{L1}$	**0.972 ± 0.006**	0.960 ± 0.001	0.941 ± 0.010	**0.886 ± 0.007**	**0.834 ± 0.010**	**0.890 ± 0.034**	0.000 ± 0.000	0.000 ± 0.000	**0.643 ± 0.153**	**0.641 ± 0.008**
$\mathcal{L}_{L2}$	0.950 ± 0.007	0.957 ± 0.009	0.939 ± 0.003	0.831 ± 0.020	0.723 ± 0.068	0.837 ± 0.051	0.888 ± 0.000	0.881 ± 0.021	0.491 ± 0.136	0.428 ± 0.196
$\mathcal{L}_{L2} \rightarrow \mathcal{L}_{L1}$	0.970 ± 0.006	**0.961 ± 0.007**	**0.941 ± 0.003**	0.869 ± 0.016	0.793 ± 0.021	0.886 ± 0.027	**0.914 ± 0.008**	**0.933 ± 0.010**	0.632 ± 0.170	0.567 ± 0.094

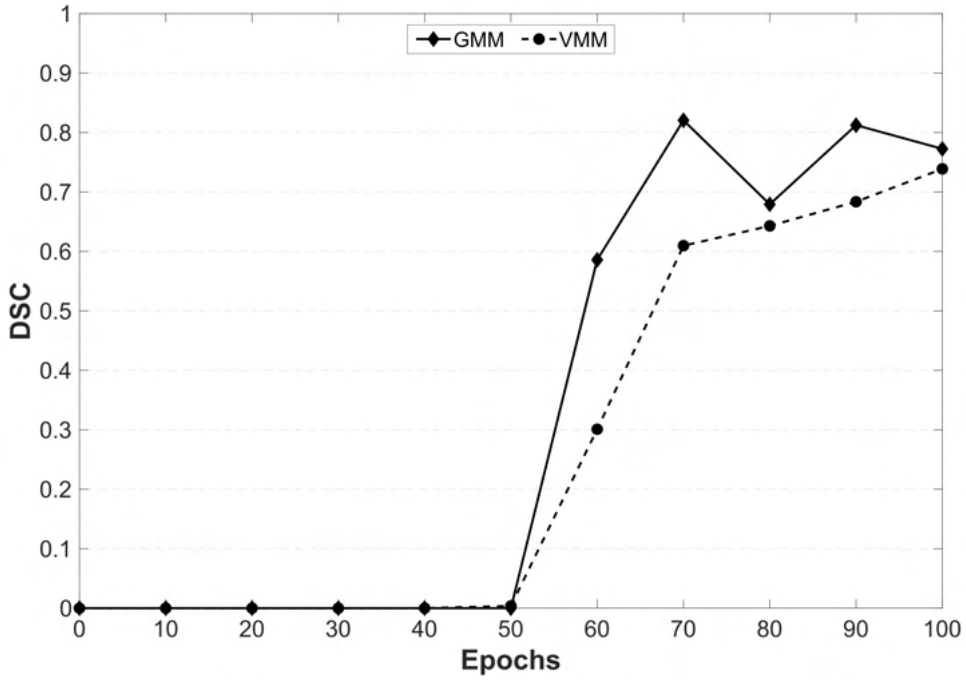

Figure 4.5: Influence of altering the loss objective during cGAN training on the segmentation performance of the medial gastrocnemius and vastus muscles. The cGAN was trained with a $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L2}$ loss objective for 50 epochs followed by a further 50 epochs training with $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$. Abbreviations: VMM - vastus medialis muscle, GMM – medial head of gastrocnemius muscle, DSC – Dice Similarity Coefficient.

4.3.4 Evaluation of the Generator Depth

The quantitative results of assessing the impact of generator network depth on the cGANs and U-Nets segmentation performances are in Table 4.6 and Table 4.7.

The cGAN with a generator down-sampling the input through nine convolutional networks achieved the highest DSC scores for tibial and patellar bone, as well as for femoral and patellar cartilage. Femoral bone and tibial cartilage were best segmented by the cGAN with five convolutions / transpose convolutions in the generator encoding / decoding parts. The medial vastus and gastrocnemius muscles, as well as ACL and PCL were best segmented by the cGAN with seven convolutions. Training the cGAN with a minimum feature channel change of 64 resulted in the highest segmentation scores for most tissues except for femoral bone, tibial cartilage and the medial vastus muscle.

The U-Net trained with nine convolutions/transpose convolutions in the networks encoding/decoding parts achieved the highest segmentation accuracies for all but one tissue (femoral cartilage), which was slightly better segmented by the U-Net with five convolutions/transpose convolutions. Training the U-Net with a minimum feature channel change of 64 resulted in the highest DSC scores for most tissues apart from patella cartilage and ACL which were segmented best by the U-Net trained with a minimum feature channel change of 128.

It is important to note for this section that increasing the number of convolutions and feature channels in the generator network substantially increases the overall number of parameters in the network and the time per epoch required to train the network (see network architectures in the Appendix for details). A considered decision between increase in learning time and significant improvement in segmentation accuracy has to be made.

Table 4.6: Results of Varying Generator Network Depth: Number of Convolutions. The influence of varying the number of convolutions during down-sampling in the generator networks of both the cGAN and U-Net was assessed. Highest network scores achieved for each tissue are in bold.
Training and testing were performed on the AMROA training and testing datasets, respectively.
Results are presented as mean ± standard deviation.
Abbreviations: F Bone – femoral bone, T Bone – tibial bone, P Bone – patellar bone, F Cartilage – femoral cartilage, T Cartilage – tibial cartilage, P Cartilage – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient

Generator Network Depth Results – Number of Convolutions during Down-Sampling										
cGAN										
Number	**F Bone**	**T Bone**	**P Bone**	**F Cartilage**	**T Cartilage**	**P Cartilage**	**VM Muscle**	**GM Muscle**	**ACL**	**PCL**
Down Convs	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**
5	**0.928 ± 0.006**	0.929 ± 0.006	0.893 ± 0.029	0.721 ± 0.029	**0.751 ± 0.039**	0.838 ± 0.042	0.049 ± 0.069	0.000 ± 0.000	0.622 ± 0.042	0.286 ± 0.189
7	0.889 ± 0.023	0.921 ± 0.026	0.928 ± 0.002	0.764 ± 0.047	0.624 ± 0.039	0.846 ± 0.057	**0.171 ± 0.226**	**0.167 ± 0.236**	**0.626 ± 0.041**	**0.289 ± 0.408**
9	0.918 ± 0.011	**0.948 ± 0.018**	**0.928 ± 0.002**	**0.812 ± 0.002**	0.748 ± 0.042	**0.863 ± 0.043**	0.113 ± 0.085	0.000 ± 0.000	0.577 ± 0.020	0.073 ± 0.103
U-Net										
Number	**F Bone**	**T Bone**	**P Bone**	**F Cartilage**	**T Cartilage**	**P Cartilage**	**VM Muscle**	**GM Muscle**	**ACL**	**PCL**
Down Convs	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**
5	0.969 ± 0.002	0.952 ± 0.016	0.919 ± 0.022	**0.887 ± 0.018**	0.823 ± 0.001	0.888 ± 0.031	0.000 ± 0.000	0.000 ± 0.000	0.631 ± 0.125	0.544 ± 0.249
7	0.964 ± 0.003	0.956 ± 0.005	0.921 ± 0.008	0.874 ± 0.032	0.787 ± 0.044	0.869 ± 0.029	0.000 ± 0.000	0.000 ± 0.000	0.539 ± 0.160	0.592 ± 0.120
9	**0.972 ± 0.006**	**0.960 ± 0.001**	**0.941 ± 0.010**	0.886 ± 0.007	**0.834 ± 0.010**	**0.890 ± 0.034**	0.000 ± 0.000	0.000 ± 0.000	**0.643 ± 0.153**	**0.641 ± 0.008**

Table 4.7: Results of Varying Generator Network Depth: Number of Minimum Feature Maps. The influence of starting with different numbers of minimum feature channel maps in the generator networks of both the cGAN and U-Net was assessed. Highest network scores achieved for each tissue are in bold.
Training and testing were performed on the AMROA training and testing datasets, respectively.
Results are presented as mean ± standard deviation.
Abbreviations: F Bone – femoral bone, T Bone – tibial bone, P Bone – patellar bone, F Cartilage – femoral cartilage, T Cartilage – tibial cartilage, P Cartilage – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient

Generator Network Depth Results – Number of Minimum Feature Channel Maps										
cGAN										
Feature Maps	**F Bone DSC**	**T Bone DSC**	**P Bone DSC**	**F Cartilage DSC**	**T Cartilage DSC**	**P Cartilage DSC**	**VM Muscle DSC**	**GM Muscle DSC**	**ACL DSC**	**PCL DSC**
16	0.774 ± 0.059	0.903 ± 0.040	0.858 ± 0.003	0.547 ± 0.236	0.473 ± 0.269	0.771 ± 0.070	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
32	0.899 ± 0.004	0.937 ± 0.001	0.875 ± 0.027	0.750 ± 0.028	0.720 ± 0.038	0.831 ± 0.030	**0.414 ± 0.260**	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
64	0.918 ± 0.011	**0.948 ± 0.018**	**0.928 ± 0.002**	**0.812 ± 0.002**	0.748 ± 0.042	**0.863 ± 0.043**	0.113 ± 0.085	0.000 ± 0.000	**0.577 ± 0.020**	**0.073 ± 0.103**
128	**0.925 ± 0.006**	0.935 ± 0.021	0.831 ± 0.032	0.805 ± 0.010	**0.773 ± 0.081**	0.784 ± 0.061	0.341 ± 0.256	0.000 ± 0.000	0.336 ± 0.219	0.011 ± 0.016
U-Net										
Feature Maps	**F Bone DSC**	**T Bone DSC**	**P Bone DSC**	**F Cartilage DSC**	**T Cartilage DSC**	**P Cartilage DSC**	**VM Muscle DSC**	**GM Muscle DSC**	**ACL DSC**	**PCL DSC**
16	0.966 ± 0.000	0.950 ± 0.021	0.912 ± 0.028	0.868 ± 0.011	0.795 ± 0.001	0.864 ± 0.028	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.202 ± 0.110
32	0.969 ± 0.006	0.946 ± 0.016	0.914 ± 0.005	0.875 ± 0.026	0.795 ± 0.051	0.878 ± 0.032	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.453 ± 0.039
64	**0.972 ± 0.006**	**0.960 ± 0.001**	**0.941 ± 0.010**	**0.886 ± 0.007**	**0.834 ± 0.010**	0.890 ± 0.034	0.000 ± 0.000	0.000 ± 0.000	0.643 ± 0.153	**0.641 ± 0.008**
128	0.968 ± 0.006	0.960 ± 0.004	0.929 ± 0.014	0.884 ± 0.022	0.823 ± 0.010	**0.897 ± 0.013**	0.000 ± 0.000	0.000 ± 0.000	**0.645 ± 0.053**	0.597 ± 0.025

4.3.5 Evaluation of PatchGAN Receptive Field Size

Figure 4.6 shows the qualitative comparison of the effect of using different patch sizes in the discriminator network, while the corresponding DSCs are listed in Table 4.8. The cGAN trained with the 1 x 1 PatchGAN (PixelGAN) achieved the highest segmentation accuracy for most tissues except for femoral and tibial cartilage and both muscle tissues, which were best segmented by the 34 x 34 PatchGAN. Increasing the receptive field size increases the number of parameters in the discriminator network and therefore may be more difficult to train.

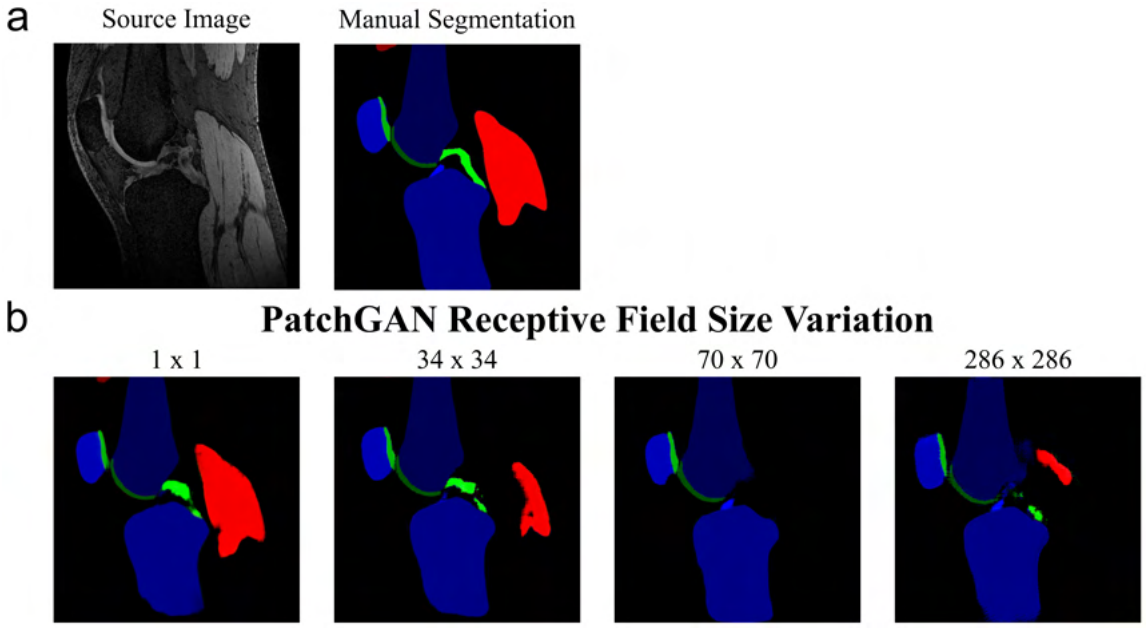

Figure 4.6: Results of PatchGAN Receptive Field Size. Assessing the influence of varying the discriminator receptive field size on segmentation performance of cGAN when trained and tested on the AMROA dataset.

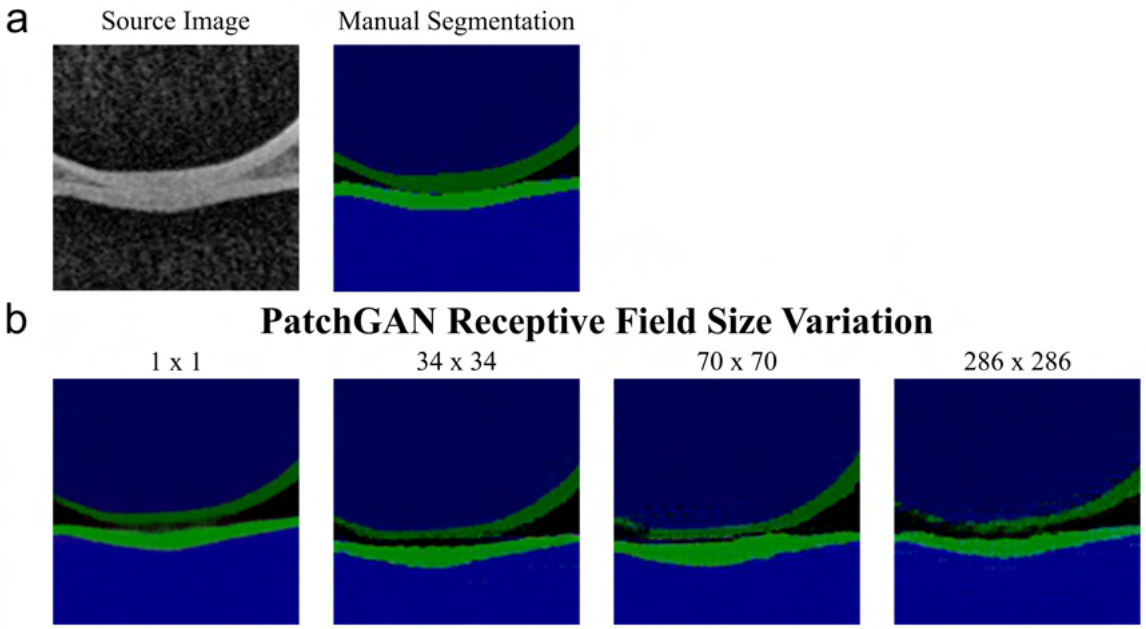

Figure 4.7: Image Artefact due to the choice of PatchGAN Receptive Field Size. Influence of discriminator receptive field size on checkerboard artefact emergence of a cGAN trained and tested on the AMROA dataset.

Additionally, as in the 'pix2pix' paper [97], we also noticed the repetitive tiling / checkerboard artefact (Figure 4.7). However, in our instance, the artefacts become more pronounced with every increase in patch size instead of the inverse tendency as seen by Isola *et al* [97]. This could be a result of us assigning the cGANs with the reverse task (image to label) compared to the one performed by Isola *et al* (label to image) [97].

Figure 4.8 depicts the loss evolution during network training of the cGAN trained with the 1 x 1 PatchGAN discriminator. The loss evolutions of the cGAN generator ($\mathcal{L}_{cGAN}$ and $\mathcal{L}_{L1}$) and discriminator ($\mathcal{L}_{real}$ and $\mathcal{L}_{fake}$) are shown in Figure 4.8A and Figure 4.8B, respectively. Figure 4.8B highlights how the Nash equilibrium was reached for the discriminator network during cGAN training.

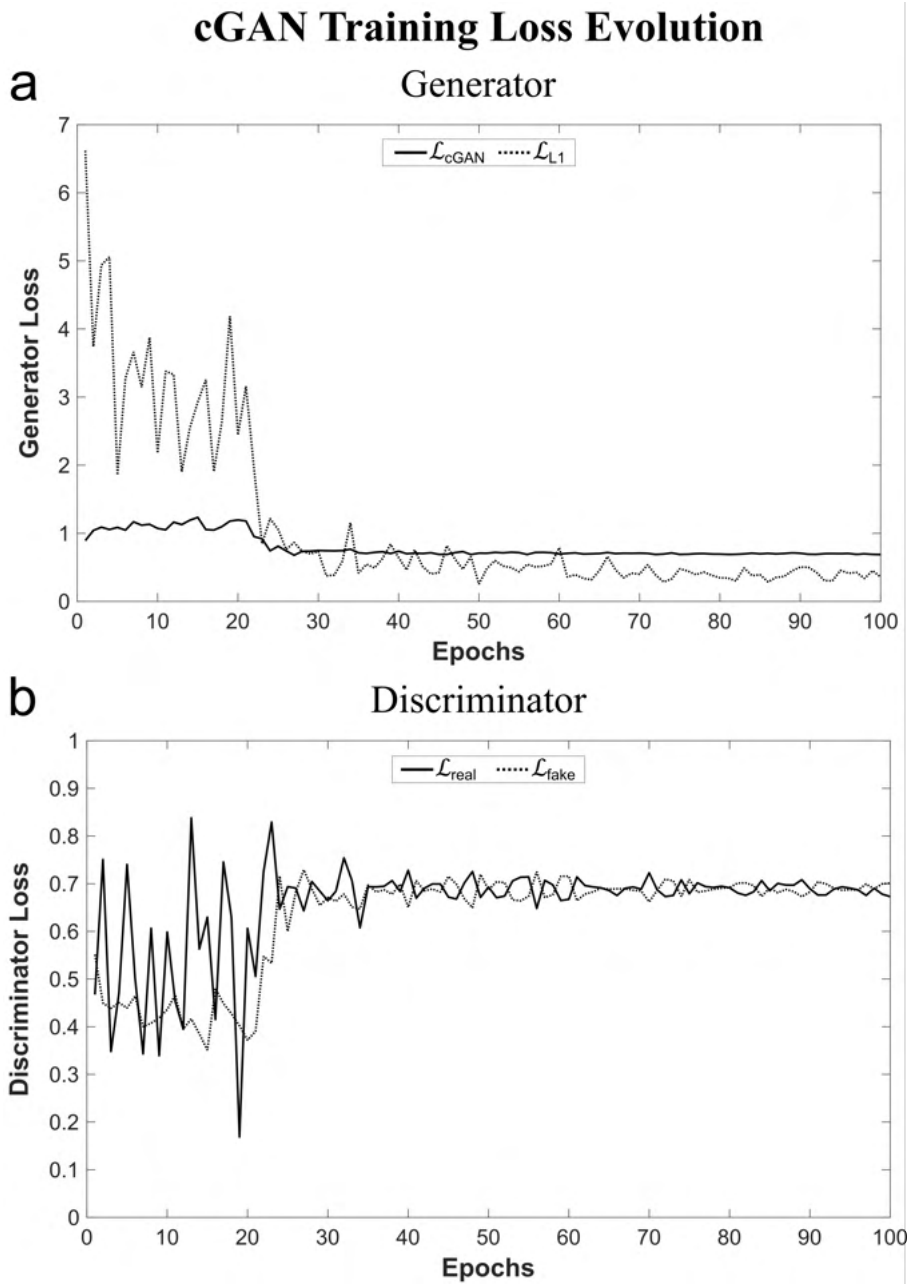

Figure 4.8: Loss Evolution during cGAN Training. The loss evolutions of the a) generator ($\mathcal{L}_{cGAN}$ and $\mathcal{L}_{L1}$) and b) discriminator ($\mathcal{L}_{real}$ and $\mathcal{L}_{fake}$) are shown for a cGAN trained with a U-Net generator and a 1 x 1 PatchGAN discriminator for 100 epochs.

Table 4.8: Results of PatchGAN Receptive Field Size. Comparison of segmentation performance of the proposed cGAN with different N x N receptive field sizes of the PatchGAN discriminator network. Highest network scores achieved for each tissue are highlighted grey and in bold. The cGANs were trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ objective with $\lambda = 100$ with training and testing being performed on the AMROA dataset. Abbreviations: FB – femoral bone, TB – tibial bone, PB – patellar bone, FC – femoral cartilage, TC – tibial cartilage, PC – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient

PatchGAN Receptive Field Size Results										
Receptive Field Size	**F Bone DSC**	**T Bone DSC**	**P Bone DSC**	**F Cartilage DSC**	**T Cartilage DSC**	**P Cartilage DSC**	**VM Muscle DSC**	**GM Muscle DSC**	**ACL DSC**	**PCL DSC**
1 x 1	**0.971 ± 0.005**	**0.953 ± 0.012**	**0.947 ± 0.007**	0.849 ± 0.046	**0.804 ± 0.024**	**0.869 ± 0.053**	0.812 ± 0.066	0.869 ± 0.069	0.618 ± 0.140	**0.613 ± 0.143**
34 x 34	0.968 ± 0.007	0.952 ± 0.015	0.941 ± 0.013	**0.849 ± 0.002**	0.795 ± 0.013	0.868 ± 0.023	**0.883 ± 0.007**	**0.876 ± 0.009**	**0.621 ± 0.096**	0.594 ± 0.118
70 x 70	0.918 ± 0.011	0.948 ± 0.018	0.928 ± 0.002	0.812 ± 0.002	0.748 ± 0.042	0.863 ± 0.043	0.113 ± 0.085	0.000 ± 0.000	0.577 ± 0.020	0.073 ± 0.103
286 x 286	0.941 ± 0.000	0.938 ± 0.008	0.920 ± 0.012	0.766 ± 0.020	0.731 ± 0.003	0.767 ± 0.049	0.702 ± 0.022	0.597 ± 0.078	0.383 ± 0.090	0.070 ± 0.022

4.3.6 Evaluation of Transfer Learning

The quantitative results of this section are presented in Table 4.9 and Table 4.10 with qualitative comparisons between single step (one dataset) and two step training (transfer learning) displayed in Figure 4.9 and Figure 4.10.

When comparing the segmentation performances of the proposed cGAN and U-Net without and with transfer learning and testing on the SKI10 testing dataset (Table 9, Figure 4.9A-C), the AMROA-pretrained / SKI10-retrained (AMROA $\rightarrow$ SKI10) U-Net showed the highest DSC scores for femoral and tibial bone and the highest boundary accuracy (i.e. smallest ASDs) for femoral bone, while the SKI10-only trained U-Net segmented the tibial bone with the highest boundary accuracy. Femoral cartilage was best segmented by the AMROA-pretrained / SKI10-retrained (AMROA $\rightarrow$ SKI10) cGAN and tibial cartilage by the SKI10-only trained cGAN.

Testing the OAI ZIB testing dataset on the proposed cGAN and U-Net without and with transfer learning (Table 9, Figure 4.9D-F), the AMROA-pretrained / OAI ZIB-retrained (AMROA $\rightarrow$ OAI ZIB) cGAN showed the highest accuracies for tibial bone and femoral cartilage, while the OAI ZIB-only trained cGAN segmented the femoral bone and tibial cartilage with the highest accuracies.

When testing the cGANs and U-Nets on the AMROA testing dataset (Table 10, Figure 4.10), the SKI10-pretrained / AMROA-retrained (SKI10 $\rightarrow$ AMROA) U-Net had the highest DSCs for femoral and tibial bone as well as the ACL. Femoral cartilage as well as patellar bone and cartilage was segmented most accurately by the OAI ZIB-pretrained / AMROA-retrained (OAI ZIB $\rightarrow$ AMROA) U-Net. The AMROA only trained U-Net showed the best segmentation accuracy for tibial cartilages. The SKI10-pretrained / AMROA-retrained (SKI10 $\rightarrow$ AMROA) cGAN provided the highest segmentation score for the vastus medialis muscle while the medial head of gastrocnemius muscle and the PCL was best segmented by the OAI ZIB-pretrained / AMROA-retrained (OAI ZIB $\rightarrow$ AMROA) cGAN. Compared to the U-Net, the cGAN could successfully segment both medial muscles which could promote a strength of the cGAN. A further note is that, although the SKI10 and OAI ZIB datasets only comprised of segmentations of femoral and tibial bone and cartilage, the cGANs and U-Nets initialised with the respective SKI10- and OAI ZIB-pretrained network weights and retrained on the AMROA dataset were able to recuperate and capture patellar, ligament and muscle tissues.

A challenge of any machine learning technique is obtaining a training set that optimises the amount of variation from the rare morphology of pathological conditions or image artefacts. The AMROA dataset was highly controlled, with the patients and imaging occurring with a single imaging protocol on a single MRI system. The images showed a clear bone-cartilage separation and enabled better cartilage segmentation scores after training than the SKI10 dataset. The OAI ZIB dataset highlights the benefits of training on a very large number of images with the cGAN and U-Net (OAI ZIB-only trained) achieving DSC ≥ 0.984 for bone and DSC ≥ 0.837 for cartilage segmentations.

The ability for the network to be used under variable conditions was simulated by using three knee datasets (AMROA, SKI10 and OAI ZIB). Even without transfer learning, the AMROA training enabled SKI10 and OAI ZIB segmentation and vice versa, albeit not with high accuracy, but nonetheless indicating the robustness of deep learning methods. Transfer learning not only improved the segmentation accuracy for some tissues of the local dataset but also enhanced the networks ability to segment the SKI10 / OIA ZIB test dataset by introducing more heterogeneity into the model. Even though the SKI10- and OAI ZIB-pretrained networks were then fine-tuned to segment the local AMROA dataset, it could segment the SKI10 and OAI ZIB

Table 4.9: Results of Transfer Learning. Comparison of segmentation performance of the proposed cGAN and U-Net without and with transfer learning and testing on the SKI10 and OAI ZIB testing dataset. Highest network scores achieved for each tissue are in bold.
SKI10/OAI ZIB → AMROA: Pretraining the network for 20 epochs on the SKI10/OAI ZIB dataset followed by network fine-tuning for 80 epochs on the AMROA dataset.
AMROA → SKI10/OAI ZIB: Pretraining the network for 20 epochs on the AMROA dataset followed by network fine-tuning for 80 epochs on the SKI10/OAI ZIB dataset.
Abbreviations: FB – femoral bone, TB – tibial bone, FC – femoral cartilage, TC – tibial cartilage, DSC - Sørensen–Dice similarity coefficient, ASSD – average surface distance, VOE – volumetric overlap error

Transfer Learning Results

SKI10 Testing

Network	Training	F Bone		T Bone		F Cartilage		T Cartilage	
		DSC	ASSD	DSC	ASSD	DSC	VOE	DSC	VOE
cGAN	AMROA	0.929 ± 0.040	3.726 ± 1.758	0.893 ± 0.069	3.368 ± 1.935	0.488 ± 0.093	67.19 ± 8.36	0.465 ± 0.114	69.01 ± 10.00
	SKI10	0.974 ± 0.013	1.445 ± 1.918	0.979 ± 0.007	0.527 ± 0.403	0.736 ± 0.058	41.49 ± 6.99	**0.684 ± 0.070**	**47.58 ± 7.98**
	SKI10 → AMROA	0.938 ± 0.039	3.229 ± 1.776	0.929 ± 0.041	2.696 ± 2.326	0.544 ± 0.077	62.23 ± 7.45	0.480 ± 0.100	67.86 ± 8.89
	AMROA → SKI10	0.974 ± 0.012	1.280 ± 1.484	0.977 ± 0.010	0.802 ± 1.139	**0.738 ± 0.059**	**41.19 ± 7.08**	0.675 ± 0.071	48.65 ± 7.94
U-Net	AMROA	0.925 ± 0.038	1.856 ± 0.997	0.907 ± 0.055	1.868 ± 1.336	0.545 ± 0.082	62.16 ± 7.62	0.462 ± 0.112	69.26 ± 9.86
	SKI10	0.973 ± 0.015	0.756 ± 0.995	0.978 ± 0.008	**0.254 ± 0.340**	0.728 ± 0.058	42.42 ± 6.88	0.674 ± 0.066	48.85 ± 7.55
	SKI10 → AMROA	0.943 ± 0.032	1.071 ± 0.682	0.936 ± 0.038	1.436 ± 1.083	0.576 ± 0.078	59.18 ± 7.86	0.456 ± 0.115	69.76 ± 9.93
	AMROA → SKI10	**0.975 ± 0.013**	**0.440 ± 0.492**	**0.979 ± 0.007**	0.258 ± 0.288	0.731 ± 0.056	42.08 ± 6.74	0.670 ± 0.070	49.19 ± 7.84

OAI ZIB Testing

Network	Training	F Bone		T Bone		F Cartilage		T Cartilage	
		DSC	ASSD	DSC	ASSD	DSC	VOE	DSC	VOE
cGAN	AMROA	0.939 ± 0.016	4.153 ± 1.962	0.914 ± 0.080	4.681 ± 3.197	0.611 ± 0.068	55.66 ± 7.10	0.601 ± 0.089	56.44 ± 9.14
	OAI ZIB	**0.985 ± 0.002**	**0.328 ± 0.123**	0.985 ± 0.003	0.293 ± 0.072	0.895 ± 0.023	18.92 ± 3.64	**0.839 ± 0.040**	**27.55 ± 5.90**
	OAI ZIB → AMROA	0.961 ± 0.009	1.786 ± 1.202	0.961 ± 0.018	4.426 ± 2.902	0.641 ± 0.071	52.41 ± 7.87	0.738 ± 0.055	41.23 ± 6.70
	AMROA → OAI ZIB	0.985 ± 0.002	0.403 ± 0.268	**0.985 ± 0.003**	**0.293 ± 0.068**	**0.897 ± 0.022**	**18.68 ± 3.57**	0.837 ± 0.042	27.82 ± 6.19
U-Net	AMROA	0.934 ± 0.015	5.424 ± 2.799	0.915 ± 0.094	6.282 ± 3.647	0.643 ± 0.065	52.26 ± 7.03	0.626 ± 0.063	54.12 ± 6.74
	OAI ZIB	0.985 ± 0.002	0.388 ± 0.169	0.984 ± 0.003	0.304 ± 0.079	0.896 ± 0.020	18.83 ± 3.19	0.837 ± 0.038	27.80 ± 5.57
	OAI ZIB → AMROA	0.966 ± 0.006	1.244 ± 0.791	0.961 ± 0.017	1.880 ± 1.133	0.734 ± 0.046	41.83 ± 5.82	0.741 ± 0.058	40.83 ± 6.97
	AMROA → OAI ZIB	0.985 ± 0.002	0.390 ± 0.361	0.985 ± 0.003	0.327 ± 0.127	0.893 ± 0.023	19.24 ± 3.64	0.838 ± 0.037	27.75 ± 5.50

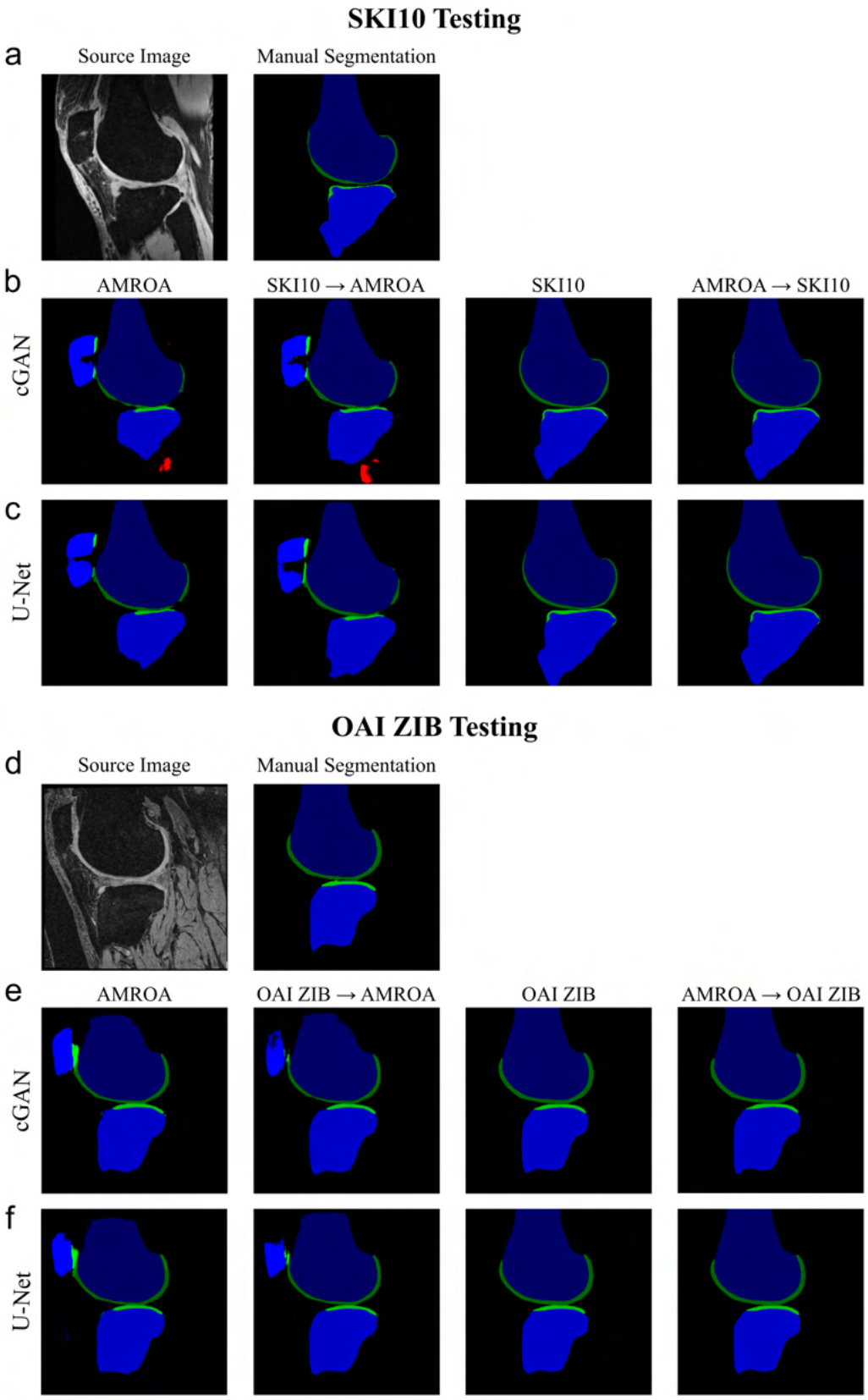

Figure 4.9: Results of Transfer Learning: SKI10 and OAI ZIB. Assessing the influence of transfer learning on segmentation performance of cGAN and U-Net when tested on the SKI10 and OAI ZIB test datasets.
SKI10 / OAI ZIB → AMROA: Pretraining the network for 20 epochs on the SKI10 / OAI ZIB training dataset followed by network fine-tuning for 80 epochs on the AMROA training dataset.
AMROA → SKI10 / OAI ZIB: Pretraining the network for 20 epochs on the AMROA training dataset followed by network fine-tuning for 80 epochs on the SKI10 / OAI ZIB training dataset.

Table 4.10: Results of Transfer Learning. Comparison of segmentation performance of the proposed cGAN and U-Net without and with transfer learning and testing on the AMROA testing dataset. Highest network scores achieved for each tissue are in bold.
SKI10/OAI ZIB → AMROA: Pretraining the network for 20 epochs on the SKI10/OAI ZIB dataset followed by network fine-tuning for 80 epochs on the AMROA dataset.
AMROA → SKI10/OAI ZIB: Pretraining the network for 20 epochs on the AMROA dataset followed by network fine-tuning for 80 epochs on the SKI10/OAI ZIB dataset.
Abbreviations: FB – femoral bone, TB – tibial bone, PB – patellar bone, FC – femoral cartilage, TC – tibial cartilage, PC – patellar cartilage, VM Muscle - vastus medialis muscle, GM Muscle – medial head of gastrocnemius medialis muscle, ACL – anterior cruciate ligament, PCL – posterior cruciate ligament, DSC - Sørensen–Dice similarity coefficient

Transfer Learning Results - AMROA Testing											
Network	**Training**	**F Bone**	**T Bone**	**P Bone**	**F Cartilage**	**T Cartilage**	**P Cartilage**	**VM Muscle**	**GM Muscle**	**ACL**	**PCL**
		DSC	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**	**DSC**
cGAN	AMROA	0.971 ± 0.005	0.953 ± 0.012	0.947 ± 0.007	0.849 ± 0.046	0.804 ± 0.024	0.869 ± 0.053	0.812 ± 0.066	0.869 ± 0.069	0.618 ± 0.140	0.613 ± 0.143
	SKI10	0.940 ± 0.024	0.947 ± 0.013		0.735 ± 0.005	0.561 ± 0.190					
	OAI ZIB	0.962 ± 0.009	0.951 ± 0.010		0.817 ± 0.032	0.790 ± 0.014					
	SKI10 → AMROA	0.970 ± 0.008	0.961 ± 0.004	0.940 ± 0.001	0.871 ± 0.029	0.774 ± 0.039	0.858 ± 0.038	**0.922 ± 0.037**	0.897 ± 0.057	0.586 ± 0.043	0.468 ± 0.186
	OAI ZIB → AMROA	0.972 ± 0.003	0.962 ± 0.001	0.947 ± 0.001	0.875 ± 0.026	0.811 ± 0.042	0.879 ± 0.022	0.908 ± 0.053	**0.909 ± 0.077**	0.664 ± 0.058	**0.652 ± 0.112**
	AMROA → SKI10	0.954 ± 0.015	0.949 ± 0.005		0.761 ± 0.025	0.544 ± 0.085					
	AMROA → OAI ZIB	0.960 ± 0.007	0.951 ± 0.012		0.821 ± 0.042	0.815 ± 0.015					
U-Net	AMROA	0.972 ± 0.006	0.960 ± 0.001	0.941 ± 0.010	0.886 ± 0.007	**0.834 ± 0.010**	0.890 ± 0.034	0.000 ± 0.000	0.000 ± 0.000	0.643 ± 0.153	0.641 ± 0.008
	SKI10	0.937 ± 0.031	0.944 ± 0.026		0.754 ± 0.009	0.637 ± 0.044					
	OAI ZIB	0.959 ± 0.003	0.953 ± 0.010		0.820 ± 0.026	0.798 ± 0.012					
	SKI10 →→ AMROA	**0.974 ± 0.003**	**0.965 ± 0.000**	0.947 ± 0.004	0.879 ± 0.012	0.815 ± 0.016	0.896 ± 0.031	0.000 ± 0.000	0.000 ± 0.000	**0.665 ± 0.114**	0.000 ± 0.000
	OAI ZIB → AMROA	0.973 ± 0.004	0.964 ± 0.005	**0.948 ± 0.005**	**0.893 ± 0.010**	0.817 ± 0.043	**0.898 ± 0.011**	0.000 ± 0.000	0.000 ± 0.000	0.648 ± 0.104	0.000 ± 0.000
	AMROA → SKI10	0.950 ± 0.031	0.959 ± 0.002		0.758 ± 0.010	0.681 ± 0.009					
	AMROA → OAI ZIB	0.962 ± 0.006	0.951 ± 0.010		0.813 ± 0.032	0.790 ± 0.039					

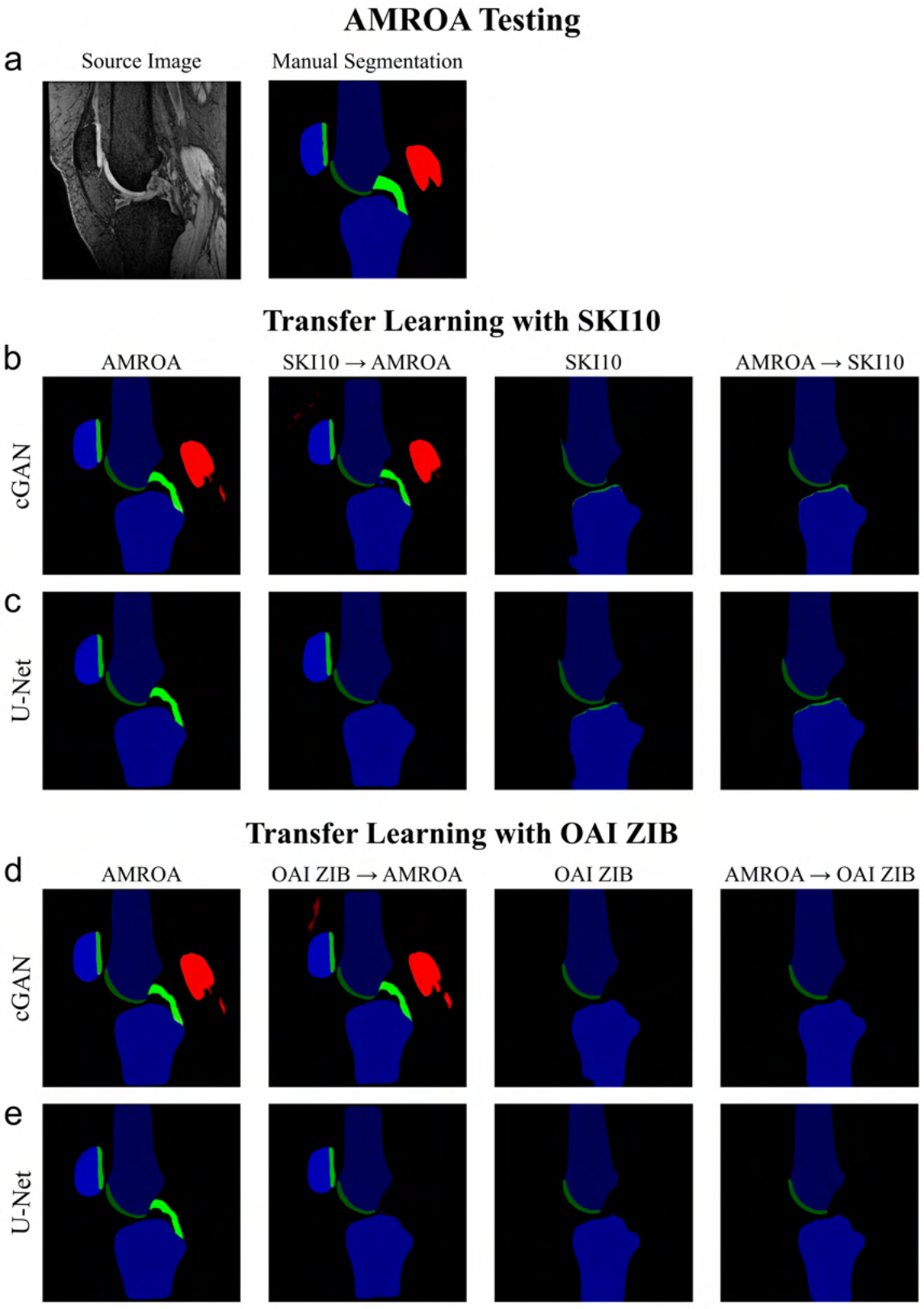

Figure 4.10: Results of Transfer Learning: AMROA. Assessing the influence of transfer learning on segmentation performance of cGAN and U-Net when tested on the AMROA test datasets.
SKI10 / OAI ZIB → AMROA: Pretraining the network for 20 epochs on the SKI10 / OAI ZIB training dataset followed by network fine-tuning for 80 epochs on the AMROA training dataset.
AMROA → SKI10 / OAI ZIB: Pretraining the network for 20 epochs on the AMROA training dataset followed by network fine-tuning for 80 epochs on the SKI10 / OAI ZIB training dataset.

testing dataset with an improved performance compared to the AMROA-only trained network without pretraining. This effect was seen for both cGANs and U-Nets.

4.3.7 AMROA: Comparison to Previous Studies

In this subsection, the results obtained for the different tissues semantically segmented in this study are compared to those of previous studies. The cGAN and U-Net achieving the highest segmentation accuracy on the AMROA dataset for each respective tissue is chosen for this purpose.

Bone

While cartilage has been traditionally studied for OA, bone shape has been under increasing investigations [153, 163]. Bone shape has been linked to radiographic OA [129, 131, 132] and associated with longitudinal pain progression [129]. Segmented bone can be used to separate out bone-specific diseases, such as osteochondral defects.

The OAI ZIB-pretrained / AMROA-retrained cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss objective (λ = 100) and a 1 x 1 PixelGAN generated segmentations of femoral (DSC = 0.972), tibial (DSC = 0.962) and patellar (DSC = 0.947) bone with the highest accuracy. The SKI10-pretrained / AMROA-retrained U-Net ($\mathcal{L}_{L1}$ loss objective) achieved slightly higher segmentation accuracies for femoral and tibial bone tissues (femoral: DSC = 0.974; tibial: DSC = 0.965) and the OAI ZIB-pretrained / AMROA-retrained U-Net for patellar bone (DSC = 0.948), compared to the cGANs. The boundaries of the images, near the top and bottom of any 2D slice, did not always segment all bone, which is where the MRI radiofrequency (RF) transmit and receive uniformity was poor due to characteristics of the MRI coil. Traditional semi-automatic approaches involving signal threshold, region-based or clustering segmentation can be similarly sensitive to image non-uniformities [141]. These non-uniformities are shown as a change in signal-to-noise or darkening of the surrounding muscle tissues (see lower regions of Figure 2). These effects from RF transmit or receive non-uniformity could be mitigated with a larger training population, as more complex modelling of data is possible. Nevertheless, segmentation of the patella achieved the lowest accuracy. The patella has the widest range of inter-subject variability when compared to the larger tibial and femoral bones. The patella bone can vary in both shape and position, shifting due to the orientation and bend of the knee. Additionally, due to its smaller volume, fewer training images are used for the patella segmentation.

The cGAN and U-Net bone segmentation scores achieved in this study are similar to those achieved by a CycleGAN method using unannotated knee MR images for femoral (DSC = 0.95 – 0.97) and tibial (DSC = 0.93 – 0.95) bone segmentation ([95], and a convolutional encoder-decoder network combined with a 3D fully connected conditional random field and simplex deformable modelling for femoral (DSC = 0.970), tibial (DSC = 0.962) and patellar (DSC = 0.898) bone segmentation [164].

Cartilage

For a long time, OA was considered a disease primarily involving variations in articular cartilage composition and morphology. Therefore, the attention was predominantly placed on the extraction of OA biomarkers from quantitative MR imaging techniques using manual or semi-manual segmentation techniques that suffer from intra- and inter-observer variability [165].

Deep learning methods can provide a fast and repeatable alternative to overcome these time-consuming and operator-dependent procedures.

The OAI ZIB-pretrained / AMROA-retrained cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss objective (λ = 100) and a 1 x 1 PixelGAN generated segmentations of femoral (DSC = 0.875), tibial (DSC = 0.811) and patellar (DSC = 0.879) cartilage with the highest accuracy from all cGAN trainings. The OAI ZIB-pretrained / AMROA-retrained U-Net ($\mathcal{L}_{L1}$ loss objective) achieved marginally higher accuracies for femoral (DSC = 0.893) and patellar (DSC = 0.898) cartilage segmentations and the AMROA-only trained U-Net ($\mathcal{L}_{L1}$ loss objective) achieved a slightly higher segmentation accuracy for tibial cartilage (DSC = 0.834) compared to the cGAN results.

The cartilage segmentation performances of both cGAN and U-Net are comparable to those attained by a 2D U-Net for femoral, tibial and patellar cartilage segmentations on T1ρ-weighted (DSC = 0.632 - 0.702) and DESS MR images (DSC = 0.767 - 0.878) [145], a CycleGAN method for femoral and tibial cartilage segmentation on PD-weighted (DSC = 0.65 - 0.66) and T2-weighted FSE images (DSC = 0.81 - 0.75) [95], as well as the recently investigated cGAN for femoral, tibial and patellar segmentation on DESS MR images (DSC = 0.843 - 0.918) [94].

Muscle

As muscle weakness and atrophy can be regarded as preceding risk factors and resulting pain-related consequences for the development and progression of OA, studying morphological changes in knee joint muscles has become increasingly important [166, 167].

The SKI10-pretrained / AMROA-retrained cGAN and the OAI ZIB-pretrained / AMROA-retrained cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss objective (λ = 100) and a 1 x 1 PixelGAN segmented the medial gastrocnemius muscle (DSC = 0.909) and medial vastus muscle (DSC = 0.922) with the highest accuracies, respectively. The U-Net trained with altered loss objective ($\mathcal{L}_{L2} \rightarrow \mathcal{L}_{L1}$) achieved the highest segmentation accuracies for both the medial gastrocnemius (DSC = 0.933) and vastus (DSC = 0.914) muscles.

Our results are comparatively lower compared to those of a semi-automatic single-atlas (DSC = 0.95 - 0.96) and fully-automatic multi-atlas (DSC = 0.91 – 0.94) based approach for medial vastus segmentation [168], and a 2D U-Net for quadriceps (DSC = 0.98) segmentation [169]. A crucial difference between these studies and ours is the plane in which segmentation was performed. While muscles are typically segmented on axial images as this provides a more straightforward task with clearer separation between different muscles, our multi-class tissue segmentation approach was performed on sagittal images. Segmenting different muscles in the sagittal plane is a demanding task, especially in areas of the calf muscles where the two-headed gastrocnemius muscle overlaps (medial and lateral) while also overlaying the soleus muscle.

Cruciate Ligament

There has been a growing interest in investigating and understanding the mechanism responsible for the post-traumatic development of OA following injury to the cruciate ligaments, especially the ACL [170–172]. Although ACL reconstruction and rehabilitation can help restore patients to normal life and previous activities, it cannot prevent the long-term risk of developing OA [173]. Accurate and repeatable segmentations of the cruciate ligaments are required when aiming at evaluating longitudinal changes in the cruciate ligaments following reconstructive surgery.

In our study, the OAI ZIB-pretrained / AMROA-retrained cGAN trained with the 1 x 1 PixelGAN and $\mathcal{L}_{cGAN}+\lambda\mathcal{L}_{L1}$ loss objective (λ = 100) achieved the highest accuracy for ACL (DSC = 0.664) and PCL segmentation (DSC = 0.652). The SKI10-pretrained / AMROA-retrained U-Net ($\mathcal{L}_{L1}$ loss objective) achieved a similar accuracy for ACL segmentation (DSC = 0.665) and the AMROA-only trained U-Net ($\mathcal{L}_{L1}$ loss objective) achieved a marginally lower accuracy for PCL segmentation (DSC = 0.641), compared to the best performing cGANs.

Lee *et al* [174] proposed a graph cut method for automatic ACL segmentation and attained a DSC score of 0.672, while Paproki *et al* [175] used a patch-based method for PCL segmentation to achieve a DSC score of 0.744. Using a 3D convolutional neural network (CNN), Mallya *et al* achieved DSC scores of 0.40 and 0.61 for ACL and PCL segmentations, respectively [176]. When combining their 3D CNN with a deformable atlas-based segmentation method, their ACL (DSC = 0.84) and PCL (0.85) segmentation accuracies increased substantially. In general, 3D networks could provide higher segmentation accuracies especially for fine structures such as the cruciate ligaments that only appear on a few 2D slices in a 3D dataset. However, 2D segmentation techniques are useful for broader applicability, as 2D imaging is often faster and currently still more clinically employed than 3D imaging.

The lower similarity scores achieved in our study compared to the other studies could arise from the use of 3D-FS SPGR images as source images during training as these are non-optimal for the segmentation of the cruciate ligaments due to their less than ideal soft tissue separation with surrounding structures and fluid. Fat-saturated proton-density-weighted fast spin echo or T2-weighted fast spin echo images are more suitable for segmentation purposes as shown by Mallya *et al* and Paproki *et al*, respectively [175, 176]. These sequences are clinically used for cruciate ligament assessment due to their dark appearance and clear separation from fluid and other surrounding tissues.

4.3.8 SKI10 and OAI ZIB: Comparison to Previous Studies

In this subsection, the segmentation results of the SKI10 and OAI ZIB datasets in this study are compared to those of previous studies. The cGAN and U-Net achieving the highest segmentation accuracy on these datasets is chosen for this purpose.

SKI10

The AMROA-pretrained / SKI10-retrained U-Net ($\mathcal{L}_{L1}$ loss objective) achieved a comparable ASSD score for femoral bone (ASSD = 0.44 mm) and an improved ASSD score for tibial bone (ASSD = 0.26 mm) to those reported by Liu *et al* and Ambellan *et al* [70, 153]. However, the segmentation accuracies for femoral (VOE $\geq$ 42.2%) and tibial (VOE $\geq$ 47.6%) cartilage achieved by our models were substantially lower.

OAI ZIB

The OAI ZIB-only trained cGAN trained with the $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss objective (λ=100) and a 1 x 1 PixelGAN generated segmentations of femoral bone (DSC = 0.985) and tibial cartilage (DSC = 0.839) with the highest accuracy. AMROA-pretrained / OAI ZIB-retrained cGAN trained with the 1 x 1 PixelGAN and $\mathcal{L}_{cGAN} + \lambda\mathcal{L}_{L1}$ loss objective (λ=100) achieved the highest accuracy for tibial bone (DSC = 0.985) and femoral cartilage (DSC = 0.897) segmentation. The ASSD of both the femoral (ASSD = 0.33 mm) and tibial (ASSD = 0.29 mm) bones were

smaller than image resolution of the OAI DESS images ($0.36 \times 0.36 \times 0.7$ mm^3). Although we achieve similar DSC scores for all tissues on the OAI ZIB dataset compared to those presented in Ambellan *et al* our ASSD scores were larger [153]. The pixel-wise error losses ($\mathcal{L}_{L1}$, $\mathcal{L}_{L2}$ and $\mathcal{L}_{SmL1}$) used to train the networks in our work were chosen to maintain an effective comparison between the cGAN and the U-Net. However, training our models with loss functions more traditionally used for segmentation purposes such as multi-class Dice similarity or cross entropy might lead to more comparable results for boundary-distance-based metrics.

4.3.9 Limitations

The network performances are depended on the accuracy of the ground truth segmentations. Inaccuracies or errors in the segmentation maps could result in a less accurate network, especially when trained on a low number of image volumes, as done in this study. Additionally, training a network on a low number of high-quality images restricts the networks applicability to only highly controlled studies with homogeneous data. Therefore, the networks trained in this study might be limited in their application in clinical settings where high image quality is not always achievable due to patient conditions and operator variabilities.

Network training on 2D MR image slices is considerably less computationally demanding than on 3D volumes. For the purposes of this study such as investigating the effects of training with different loss objectives and cGAN discriminator networks, it was sufficient to train on 2D images. Nevertheless, the segmentation of small knee joint structures, such as the cruciate ligaments, could benefit from 3D networks that should add spatial continuity along the slice dimension.

Furthermore, the segmentation results presented in this study are from standalone networks without further processing within a pipeline. Therefore, the obtained results, especially for cartilage segmentation, are not comparable to those from current state-of-the-art pipeline methods such as described by Liu *et al* [70] and Ambellan *et al* [153] that initially perform automated segmentation using a CNN followed by further refinement using deformable or statistical shape models, respectively.

Lastly, additional investigations into varying the network architectures and optimisation strategies are warranted, with ever more loss functions as well as layer combination and optimisation strategies continuously being developed.

4.4 Conclusion

This work demonstrated the usage of a cGAN, using a U-Net generator with a PatchGAN discriminator, for the purpose of automatically segmenting multiple knee joint tissues on MR images. While DSC > 0.9 were achieved for all segmented bone structures and DSC > 0.75 for cartilage and muscle tissues, DSC of only $\approxeq$0.64 were achieved for cruciate ligament segmentations. Nevertheless, this segmentation performance was attained despite the low number of subjects (N=8) for training on the local dataset. Although the U-Net outperformed the cGAN in most knee joint tissue segmentations, this study provides an optimal platform for future technical developments for utilising cGANs for segmentation tasks. By enabling automated and simultaneous segmentation of multiple tissues we hope to increase the accuracy and time efficiency for evaluating joint health in osteoarthritis.

4.5 Appendix

Generator: The encoding part of the generator network consists of the repeated application of nine 4x4 convolutions with stride 2, down-sampling the input by a factor of 2 at each layer. Each convolution is followed by a batch normalisation layer (except the first layer) and a leaky rectified linear unit (leaky ReLU) with slope 0.2. During the first encoding step the number of feature channels is changed from 3 to 64. At the subsequent three encoding steps, the number of feature channels is doubled (64 – 512), while the following five are kept at 512. In the ensuing decoding part, the input is repeatedly up-sampled by a factor of 2 by nine 4x4 transpose convolutional layers with stride 2 and additional skip connections (concatenations) between each layer i and 9-i, changing the number of feature channels at each step. The first four decoder convolutions are followed by batch normalisation, dropout (50%) and a ReLU. The next four decoder convolutions are followed by batch normalisation and a ReLU without dropout. After the final layer a convolution followed by a Tanh activation layer is applied to generate the segmentation map.
Total number of parameters: 66.999 M

Training time (s/epoch):	AMROA:	135 (cGAN with 1x1 PixelGAN)
		130 (cGAN with 70x70 PatchGAN)
		100 (U-Net)
	SKI10:	380 (cGAN with 1x1 PixelGAN)
		210 (U-Net)
	OAI ZIB:	2710 (cGAN with 1x1 PixelGAN)
		1530 (U-Net)

Generator with five convolutions in encoder/decoder: In this generator network, the encoding part consists of the repeated application of five 4x4 convolutions with stride 2, down-sampling the input by a factor of 2 at each layer. In the ensuing decoding part, the input is repeatedly up-sampled by a factor of 2 by five 4x4 transpose convolutional layers with stride 2 and additional skip connections between each layer i and 5-i.
Total number of parameters: 16.659 M

Training time (s/epoch):	AMROA:	110 (cGAN with 70x70 PatchGAN)
		90 (U-Net)

Generator with seven convolutions in encoder/decoder: The encoding part consists of the repeated application of seven 4x4 convolutions with stride 2, down-sampling the input by a factor of 2 at each layer. In the subsequent decoding part, the input is repeatedly up-sampled by a factor of 2 by seven 4x4 transpose convolutional layers with stride 2 and additional skip connections between each layer i and 7-i.
Total number of parameters: 41.829 M

Generator with 16 as minimum number of feature channels: In this network, the number of feature channels is changed from 3 to 16 during the first encoding step. During the following

Training time (s/epoch):	AMROA:	120 (cGAN with 70x70 PatchGAN)
		100 (U-Net)

three encoding steps, the number of feature channels is doubled (16 – 128), while the subsequent five are kept at 128. Total number of parameters: 4.191 M

Training time (s/epoch):	AMROA:	105 (cGAN with 70x70 PatchGAN)
		70 (U-Net)

Generator with 32 as minimum number of feature channels: The number of feature channels is changed from 3 to 32 during the first encoding step. In the following three encoding steps, the number of feature channels is doubled (32 – 256), while the subsequent five are kept at 256. Total number of parameters: 16.755 M

Training time (s/epoch):	AMROA:	100 (cGAN with 70x70 PatchGAN)
		75 (U-Net)

Generator with 128 as minimum number of feature channels: In the first encoding step the number of feature channels is changed from 3 to 128. In the following three encoding steps, the number of feature channels is doubled (128– 1024), while the subsequent five are kept at 1024. Total number of parameters: 267.953 M

Training time (s/epoch):	AMROA:	245 (cGAN with 70x70 PatchGAN)
		220 (U-Net)

Discriminator:
70 x 70 PatchGAN: The discriminator network repeatedly down-samples the input by applying three 4x4 convolutions with stride 2 followed by two 4x4 convolutions with stride 1. Each convolution during down-sampling is followed by a batch normalisation layer (except the first and last layer) and a leaky ReLU (slope 0.2) (except for the last layer). The number of feature channels are doubled (64 – 512) during the first four convolutional steps. The final convolutional layer is proceeded by a Sigmoid activation layer. Total number of parameters: 2.769 M

1 x 1 PatchGAN (PixelGAN): This PixelGAN discriminator network applies three 1 x 1 convolutions with stride 1, where the first convolution is followed by a leaky ReLU (slope 0.2) , the second convolution by a batch normalisation layer and a leaky ReLU (slope 0.2) and the final convolution by a Sigmoid activation function. The number of feature channels are doubled (64 – 128) during the first two convolutions. Total number of parameters: 0.009 M

34 x 34 PatchGAN: This network repetitively down-samples the input by using two 4x4 convolutions with stride 2 followed by two 4x4 convolutions with stride 1. Each convolution is followed by a batch normalisation layer (except the first and last layer) and a leaky ReLU

(slope 0.2) (except for the last layer). The number of feature channels are doubled (64 – 256) during the first three convolutional steps. The final layer is ensued by a Sigmoid activation layer. Total number of parameters: 0.666 M

286 x 286 PatchGAN: This discriminator network consists of eight convolutional layers with 4x4 spatial filters. The first 6 convolutions have stride 2 while the last two have stride 1. Each convolutional layer is followed by a batch normalisation layer (except the first and last layer) and a leaky ReLU (slope 0.2) (except for the last layer). The number of feature channels are doubled (64 – 512) during the first four convolutions and kept at 512 for the ensuing layers. A Sigmoid activation layer succeeds the final convolution. Total number of parameters: 11.159 M

Chapter 5

Automated Segmentation of Knee MRI Data with Convolutional Neural Networks for Three-Dimensional Surface-Based Analysis of Cartilage Morphology and Composition

Submitted for publication.

5.1 Introduction

Osteoarthritis (OA) is the most common functionally disabling joint disorder. It is characterised by the progressive deterioration of articular cartilage, subchondral bone and other tissues of diarthrotic joints [177, 178]. A full understanding of all effects that contribute to OA development is lacking, hindering the development of effective interventions at early stages of disease. Quantitative magnetic resonance imaging (qMRI) methods can assist in the non-invasive detection and quantification of morphological and compositional changes present in diseases such as OA. Although several promising techniques have been validated, to date, no qMRI method has yet been regulatorily qualified as a sensitive and reliable disease biomarker . Potential qMRI biomarkers include measurement of cartilage volume and thickness, as well as the spin lattice relaxation time in the rotating frame ($T_{1\rho}$) and the transverse relaxation time (T_2) [37, 39, 40]. However, their clinical translation has been affected by the laborious post-processing required which almost always includes some form of image segmentation for detailed region- or tissue-specific analysis, limiting their use to primarily early-phase clinical trials [3, 42, 108].

Traditionally, compartmental measurements of cartilage morphology and composition have been performed in studies aimed at determining biomarkers for early OA detection and progression. Nevertheless, such measurements over large regions-of-interest could mask heterogeneous focal changes and are prone to inter- and intra-observer errors [41, 165]. More recent studies have advanced to perform analysis on multiple smaller cartilage subregions and layers aimed at avoiding the masking of important focal variations, however measurements extracted from these sub-regional compartments are even more prone to observer error [41]. As a result, cartilage surface-based analysis techniques have progressively gained more interest over recent

years as these methods can determine and visualise the heterogeneity and bidirectionality of morphological and compositional changes occurring during OA progression [3, 172, 179]. A surface-based method termed 3D cartilage surface mapping (3D-CaSM) was recently described and validated for analysing cartilage thickness and composition on MRI [3]. However, its utility for analysing large imaging cohorts is limited by requiring manually segmented bone-cartilage structures from which cartilage patches are determined and correspondent cartilage thickness maps calculated.

While manual tissue segmentation continues to be the gold-standard method for analysing qMRI data, it is very time-consuming. Therefore, interest has grown for the development of consistent automated methods for multi-tissue segmentation of MR images [165, 180]. Prior to the recent advent of machine learning, most proposed methods have been shape-, region- or atlas-based which require a priori knowledge of the image structures. Developments in deep learning (DL) using convolutional neural networks have shown great promise in overcoming the repetitive and laborious nature of manual tissue segmentation. The convolutional encoder-decoder network U-Net [93], and its 3-dimensional (3D) analogues, 3D U-Net [181] and V-Net [182], are currently regarded as state-of-the-art methods, showing high segmentation accuracy in musculoskeletal segmentation tasks [70, 94, 145, 183].

Increasingly, studies are investigating the utility of DL methods for the segmentation of musculoskeletal MR images. However very few have assessed the efficacy of the segmentations for extracting accurate qMRI values. In a study by Paproki *et al*, measurements of tissue volume and T_2 relaxation times, extracted from automated posterior cruciate ligament segmentations using a patch-based method, achieved high correlations with those from manual segmentations [175]. Similarly, Norman *et al* achieved high correlations between manual and automatic quantifications of cartilage and menisci morphology (volume, thickness) and composition ($T_{1\rho}$, T_2) using a U-Net [145]. Liu *et al* showed no significant difference between the T_2 relaxation times extracted from manual and automated segmentations using SegNet [184] in combination with a 3D deformable model of femoral, tibial and patellar cartilage [70]. A recent study by Wirth *et al* evaluated the use of U-Net generated segmentations of femorotibial cartilage from two different MR image contrast for deriving accurate and longitudinally reproducible measures of cartilage morphology such as thickness, volume and surface area [185]. However, these studies have used compartmental rather than surface-based analysis which may have masked the heterogeneity in segmentation accuracy for different cartilage regions.

The purpose of this study was to analyse the usage of 2D and 3D U-Net generated segmentations of knee MR images within the 3D-CaSM method for extracting accurate regional measurements of cartilage morphology and relaxometry. We investigate if these networks can provide a fast and accurate method for the segmentation of knee MRIs that overcomes the limitations of manual segmentation. By comparing the extracted morphological and compositional cartilage measures from manual and automated segmentations we expect to be able to identify those focal regions where the networks segmentations deviate from manual.

5.2 Methods

5.2.1 Image Datasets

The cross-sectional MRI dataset used was OAI ZIB consisting of multi-class tissue segmentations of femoral and tibial cartilage and bone from 507 patients from the publicly available Os-

teoarthritis Initiative (OAI) baseline dataset [153, 154]. All images were acquired on Siemens 3T Trio systems using a 3D double echo steady state (DESS) sequence with water excitation. After automatically generating outlines of femoral and tibial bone and cartilage using a statistical shape model, manual adjustments were performed by experts at Zuse Institute Berlin (ZIB) [138]. The OAI ZIB cohort consisted of patients with various degrees of OA progression (Kellgren-Lawrence (KL) 0 – 4), with a greater amount having moderate-to-severe OA (KL $\geq$ 3). Detailed characteristics of the OAI ZIB cohort are in Ambellan et al [153].

The MRI dataset used was OAI ZIB consisting of multi-class tissue segmentations of femoral and tibial cartilage and bone from 507 patients from the publicly available Osteoarthritis Initiative (OAI) dataset [153, 154]. All images were acquired on Siemens 3T Trio systems using a 3D double echo steady state (DESS) sequence with water excitation. After automatically generating outlines of femoral and tibial bone and cartilage using a statistical shape mode, manual adjustments were performed by experts at Zuse Institute Berlin (ZIB) [138]. The OAI ZIB cohort consisted of patients with various degrees of OA progression (Kellgren-Lawrence (KL) 0 – 4), with a greater amount having moderate-to-severe OA (KL $\geq$ 3). Detailed characteristics of the OAI ZIB cohort are in Ambellan *et al* [153].

OAI T_2 Mapping

T_2-weighted images from the OAI dataset were acquired with a sagittal 2D multi-slice, multi-echo (MSME) spin echo (SE) sequence. Sequence parameters were as followed: acquisition time = 10.6 min; FOV = 120 mm; matrix = 384 x 269 interpolated to 384 x 384; slice thickness = 3 mm; approximately 27 slices per TE; TR = 2700 ms; TEs = 10, 20, 30, 40, 50, 60, 70 ms; bandwidth = 250 Hz/pixel.

For the analysis of cartilage T_2 relaxation times , all T_2-weighted images were rigidly registered to the 3D DESS images using the Elastix registration toolbox before calculating the quantitative T_2 maps.

T_2 maps were calculated using a log-linearised least-squares algorithm to fit a mono-exponential decay function to the signal intensities:

$$S = S_0 \cdot e^{-\frac{TE}{T_2}}. \tag{5.2.1}$$

T_2 relaxation times > 100ms in T_2 maps were excluded from analysis to avoid including partial volume artefacts with synovial fluid [82, 117, 186].

5.2.2 2D and 3D U-Net Model Specifications

The 2D U-Net adapted a previously described model used for multi-tissue segmentation of knee MR images [183]. To overcome the memory constraints when training the 3D U-Net, while maintaining an effective comparison between both networks, we modified the input layer of the U-Net to an input size of 256 x 256.

The 3D U-Net used a similar architecture to the 2D U-Net, except that all 2D operations were replaced with their 3D counterparts. The detailed architectures of both networks are shown in Figure 5.1.

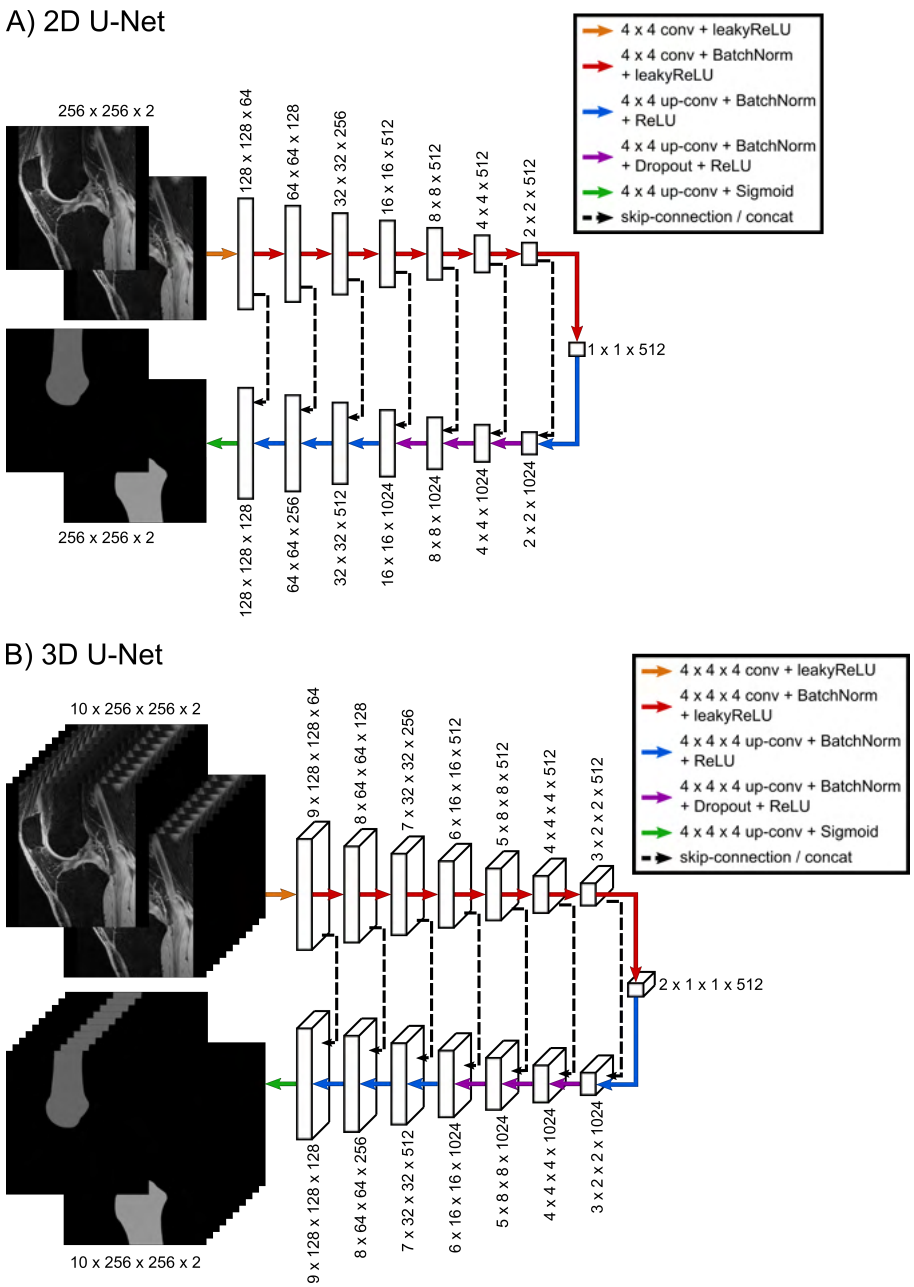

Figure 5.1: A) The 2D U-Net architecture. The encoding part of the network consists of the repeated application of eight 4x4 convolutions with stride 2 and padding 1, downsampling the input by a factor of 2 at each layer. With exception of the first layer, each convolution is followed by a batch normalisation layer (BatchNorm) and a leaky rectified linear unit (leaky ReLU) with slope 0.2. After the first encoding step, the number of feature channels is doubled during the next three convolutions (64 – 512), while the ensuring four are kept at 512. In the subsequent decoding part, the input is repeatedly up sampled by a factor of 2 by eight 4x4 convolutional layers with stride 2 and padding 1, and additional skip connections (concatenations) between each major layer, changing the number of feature channels at each step. The first transpose convolution is followed by BatchNorm and ReLU without dropout while the succeeding three transpose convolutions are followed by BatchNorm, dropout (50%) and a ReLU. The next three decoder up-convolutions are followed by BatchNorm and a ReLU without dropout. The final up-convolution is only followed by a Sigmoid activation layer to generate the label map. Total number of parameters: 66.999 M. **B)** The 3D U-Net architecture is very similar to that of the 2D U-Net in which all 2D operations are replaced by their 3D counterparts, i.e. 4x4x4 convolutions with 1x2x2 stride and 1x1x1 padding as well as 3D BatchNorm. Total number of parameters: 217.614 M

Table 5.1: Summary of testing dataset used to evaluate networks and to perform quantitative cartilage surface analysis.

Testing	Number of Subjects	20
	Sex [Female / Male]	12 / 8
	Age (years)	66.8 ± 9.0
	Subjects per KL Grade [0 / 1 / 2 / 3 / 4]	2 / 4 / 6 / 6 / 2

5.2.3 Data Preparation and Network Training

The MRIs and image masks of the OAI ZIB data were converted from their respective DICOM and ITK MetaImage Header (MHD) formats [156, 187] to a MAT-file format (binary MATLAB files) before training. Each major structure was stored in a separate channel in the segmentation map. To allow a simple downstream integration within the 3D cartilage surface mapping (3D-CaSM) method described below, the OAI ZIB femoral bone and cartilage segmentations were combined into one structure, given a value code of 1 and stored in the first channel. Similarly, the tibial bone-cartilage combination was given a value code of 1 and stored in the second channel of the volume. The two-channel input to the networks had identical DESS image slices stored in both channels. While the 2D U-Nets were trained on individual slices, for 3D network training, each dataset was split into 256 x 256 x 10 sub-volumes of overlapping slices. For instance, a single OAI 3D DESS dataset consisting of 160 slices was divided into 31 sub-volumes with each sub-volume overlapping the previous by 5 slices (except for the first and last sub-volumes, that started at slice 1 and ended at slice 160, respectively). We divided the OAI ZIB dataset into 467 cases for training, 20 cases for validation and 20 cases for testing, on which the 3D-CaSM analysis would be performed. The validation and testing sets consisted of patients covering all degrees of OA progression (KL 0 – 4), with the majority having minimal-to-moderate OA (KL 1-3). Characteristics of the testing set cohort are in Table 5.1.

Both networks were implemented using PyTorch (Torch v1.0.1) and all training was performed on a Nvidia P6000 GPU card (3840 CUDA cores, 24 GB GDDR5X). All models were trained for 100 epochs with the Adam solver [188] being used for network optimisation with a learning rate 0.0002 and momentum parameters, γ_1 = 0.5, γ_2 = 0.999. To avoid overfitting, we employed early stopping once the networks performances did not improve over 20 epochs assessed with the validation set. Batch sizes of 50 and 5 were used for 2D and 3D network training, respectively. We introduced random noise (jitter) during training by using bi-cubic interpolation to resize the input DESS images from their original size (384 x 384) to 286 x 286 and then randomly cropping the images to 256 x 256 to resemble the input layer size of the networks.

We evaluated the 2D and 3D U-Net performances by training with different loss functions. We trained the networks with binary cross entropy loss (BCE), $Dice$ loss, as well as a weighted combination of the cross entropy (CE) and $Dice$ loss within the $Combo$ loss ($\beta \cdot CE - (1 - \beta) \cdot Dice$) [189]. The weighting hyperparameter β between the CE and $Dice$ loss was altered to vary the balance between the two losses. We investigated values for β = 0.25, 0.5 and 0.75. The segmentations from the best-scoring 2D and 3D networks were chosen to perform the 3D-CaSM analysis described below.

5.2.4 Evaluation Metrics

After network training, a tissue-specific Boolean mask was created on the predicted test images.

We evaluated the network segmentation performances with the widely used Sørensen–Dice Similarity Coefficient (DSC) [89, 161], Volumetric Overlap Error (VOE) and the boundary distance-based metric Average Symmetric Surfaces Distance ($ASSD$). The DSC ranges between 0 and 1, higher DSC values representing greater overlap between the manual (X) and network generated (Y) segmentations. The DSC is given as

$$DSC = \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{5.2.2}$$

for Boolean metrics. The VOE also ranges between 0 and 1, however with small values for VOE expressing greater segmentation accuracy. It can be calculated as

$$VOE = 1 - \frac{|X \cap Y|}{|X \cup Y|} \tag{5.2.3}$$

The $ASSD$ calculates the average of all distances from each pixel on the boundary of the manual segmentation X to the boundary of the automated segmentation Y and vice versa. It is expressed in mm and is defined as

$$ASSD = \frac{1}{N_X + N_Y}\left(\sum_{i=1}^{N_X} D_X(y) + \sum_{i=1}^{N_Y} D_Y(x)\right) \tag{5.2.4}$$

where $D_X(y) = \min_{x \in X} \|y - x\|$ and $D_Y(x) = \min_{y \in Y} \|x - y\|$.

5.2.5 Postprocessing and Cartilage Surface-based Analysis

3D Cartilage Surface Mapping

Contours from the 2D and 3D U-Net generated masks of femoral and tibial bone-cartilage structures were extracted and converted into tissue-specific polygons within Python (v3.7.6, Python Software Foundation, Wilmington, Delaware, United States). Detailed surface-based analysis (3D Cartilage Surface Mapping, 3D-CaSM) of the contours generated from manual and network-automated segmentations of femoral and tibial bone-cartilage structures was performed on the OAI DESS images using the StradView software v6.1 (freely available at http://mi.eng.cam.ac.uk/Main/StradView/) [3, 186]. The full 3D-CaSM pipeline is illustrated in Figure 2. The femoral and tibial bone-cartilage contours are first used to generate 3D triangulated mesh surface objects (Figure 5.2, Step 1). By displaying the signal intensities along surface normal towards the inside of the mesh objects onto the objects surface, the cartilage surfaces can be visualised. Triangulated femoral, lateral tibial and medial tibial cartilage surface patches can then be manually extracted, taking approximately two minutes for all three cartilage surfaces (Figure 5.2, Step 2). These patches are then used to calculate the cartilage thickness at each surface vertex by generating inner and outer cartilage surfaces and sampling the length of the perpendicular line between the two surfaces (Figure 5.2, Step 3). The determined thickness measurements can then be displayed onto the surface (Figure 5.2, Step 4).

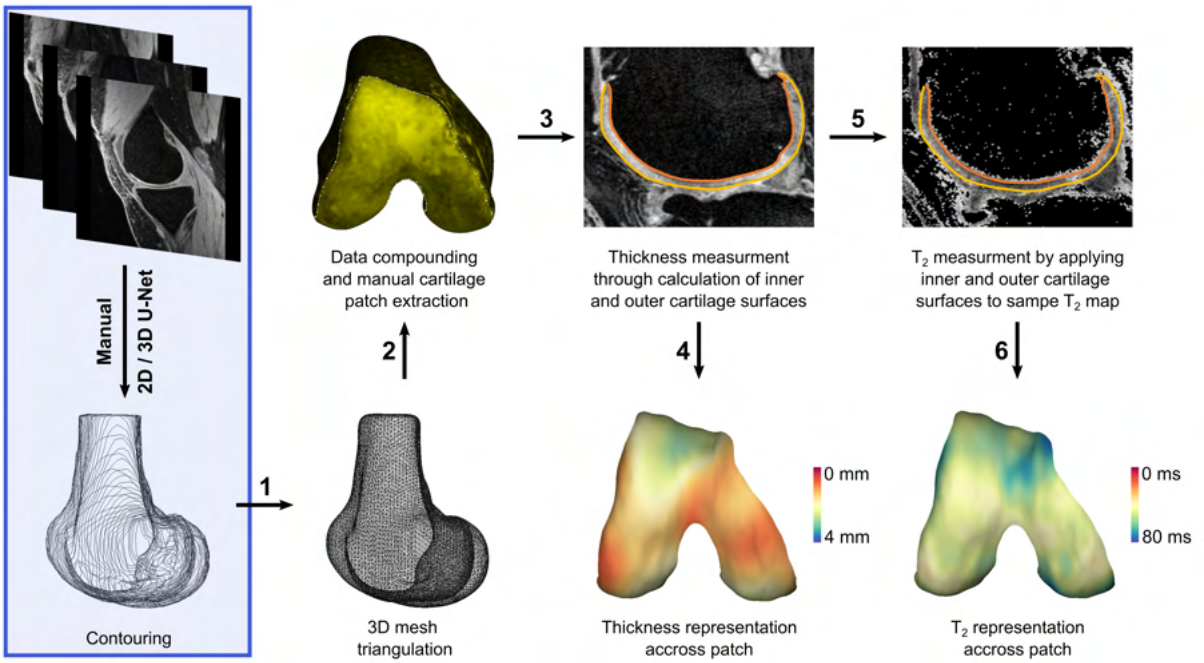

Figure 5.2: 3D Cartilage Surface Mapping (3D-CaSM) pipeline used for quantitative measurements of cartilage thickness and T_2 relaxation time determined from manual and network-generated segmentations. Femoral cartilage measurements used for demonstration purposes; same process used for tibial cartilages. Following the conversion of the manual and automated segmentation maps into tissue-specific polygon contours, 3D triangulated surface mesh objects were generated from the bone-cartilage structures (step 1). Through data compounding (displaying mean image intensities within 10 pixels from the 3D object onto its surface), cartilage patches were identified and manually extracted (step 2). Inner and outer cartilage surfaces were determined by calculating the cartilage thickness across the patch (step 3). These vertex-wise thickness measurements can then be displayed onto the cartilage surface patch (step 4). The same inner and outer cartilage surfaces can then be used to sample the registered T_2 maps and obtain vertex-wise measurements of cartilage T_2 relaxation times (steps 5 and 6). After registering individual patches to a canonical surface, we can visually compare the measurements obtained from manual and network-generated contours and calculate vertex-wise thickness and T_2 differences (step not shown). The focus of this study is highlighted in blue in which we compare the 3D-CaSM pipeline using contours from manual segmentations with the 3D-CaSM pipeline using contours determined from automated segmentations. The successive steps are kept identical for an effective comparison.

Analysis of Cartilage T_2

Following the generation of inner and outer cartilage surfaces through the thickness measurement procedure using the 3D-CaSM method, these surfaces were used to analyse the registered quantitative T_2 maps. At each triangulated surface vertex, the T_2 relaxation time values along a perpendicular line (surface normal) between inner and outer surface were sampled and averaged (Figure 5.2, Steps 5 and 6).

Cohort Analysis

For an effective comparison, each participants unique triangulated femoral, medial tibial and lateral tibial cartilage surface mesh was registered to a canonical (average) femoral, medial tibial and lateral tibial mesh. Subsequently, the individual thickness and T_2 data were mapped onto the canonical surface. Both surface registration and mapping of the individual surface data to the canonical surfaces was performed using the freely available wxRegSurf software v20 (http://mi.eng.cam.ac.uk/ ahg/wxRegSurf). Since all data were registered to a canonical

surface, vertex-wise cohort-averaged analysis was performed by calculating the root-mean-squared (RMS) thickness and T_2 at each canonical surface vertex from the individual cartilage thickness and T_2 measurements of the 20 OAI ZIB testing datasets.

5.2.6 Statistical Analysis

Two-sample t-tests were used to determine whether differences in segmentation scores between 2D U-Nets and 3D U-Nets trained with equivalent loss functions were significant. For all t-test analyses, a significant level of 0.05 was used.

The RMS thickness and T_2 measurements at each vertex from cartilage surfaces extracted from manual segmentations were compared to the matching vertex measurements on the corresponding cartilage surfaces extracted from automated segmentations. Scatterplots and Pearson correlation coefficients were determined to evaluate any related differences. Bland-Altman analysis was performed to determine the mean bias and the 95% limits of agreement between manual and automated RMS thickness and T_2 measurements.

5.3 Results

5.3.1 Network Training and Testing

Training of the 2D and 3D U-Nets on the OAI ZIB dataset required approximately 500 s/epoch and 2500 s/epoch, respectively. Segmenting an entire OAI 3D DESS volume took approximately 3.2 s with 2D U-Net and 4.6 s with 3D U-Net.

5.3.2 Segmentation Performance

Average scores on the OAI ZIB testing dataset from all networks trained with different loss objectives are shown in Table 5.2. Both 2D and 3D networks achieve high segmentation performance for femoral and tibial bone-cartilage structures compared to manual segmentations. All segmentation scores between 2D and 3D U-Nets trained with the same loss functions differed significantly ($p < 0.05$). The 2D U-Net and the 3D U-Net trained with $Combo$ loss (β = 0.25) achieved the highest segmentation performance for the femoral bone-cartilage structure. The 2D U-Net trained with $Combo$ loss (β = 0.75) and the 3D U-Net trained with $Dice$ loss-only achieved the highest segmentation performance for the tibial bone-cartilage structure. The 3D U-Net achieved higher overall segmentation performance compared to the 2D U-Net for both femoral (DSC: 0.980 vs 0.971; $ASSD$: 0.314 mm vs 0.543 mm) and tibial (DSC: 0.982 vs 0.974; $ASSD$: 0.282 mm vs 0.412 mm) bone-cartilage structures. The segmentations from the 2D U-Net and the 3D U-Net trained with $Combo$ loss (β = 0.25) were used to perform the 3D-CaSM.

5.3.3 Cartilage Surface-based Analysis

Results from the Pearson correlation and Bland-Altman analysis of thickness and T_2 measurements are listed in Table 5.3 as well as visualised in Figure 5.3 and Figure 5.4 for cartilage thickness, and Figure 5.6 and Figure 5.7 for cartilage T_2 measurements. Bland-Altman analysis

Table 5.2: Comparison of segmentation performance of the proposed 2D and 3D U-Net when trained and tested on the OAI ZIB dataset using different loss objective functions. Data are means ± standard deviations. The asterisk (*) indicates that all differences in segmentation scores between the 2D U-Nets and the corresponding 3D U-Nets of the same loss function were statistically significant ($p < 0.05$). Abbreviations: ASSD, average symmetric surfaces distance; BCE, binary cross entropy; β, weighting hyperparameter within Combo loss; DSC, Sørensen–Dice similarity coefficient; VOE, volumetric overlap error.

Network	**Loss**	β	**Femur**			**Tibia**		
			DSC	**VOE (%)**	**ASSD (mm)**	**DSC**	**VOE (%)**	**ASSD (mm)**
2D U-Net	BCE		0.966 ± 0.008	6.598 ± 1.568	0.701 ± 0.232	0.974 ± 0.004	5.143 ± 0.687	0.449 ± 0.093
	Dice		0.968 ± 0.007	6.118 ± 1.295	1.030 ± 0.717	0.968 ± 0.005	6.166 ± 0.906	0.548 ± 0.109
	Combo	0.25	**0.971 ± 0.006**	**5.662 ± 1.150**	**0.543 ± 0.164**	0.974 ± 0.004	5.129 ± 0.685	0.437 ± 0.068
		0.5	0.966 ± 0.008	6.522 ± 1.446	0.779 ± 0.335	0.972 ± 0.004	5.512 ± 0.706	0.459 ± 0.060
		0.75	0.969 ± 0.007	6.070 ± 1.258	0.581 ± 0.168	**0.974 ± 0.003**	**5.044 ± 0.646**	**0.412 ± 0.055**
3D U-Net*	BCE		0.978 ± 0.004	4.379 ± 0.840	0.353 ± 0.058	0.980 ± 0.003	3.895 ± 0.506	0.305 ± 0.039
	Dice		0.979 ± 0.004	4.069 ± 0.723	0.327 ± 0.052	**0.982 ± 0.002**	**3.621 ± 0.465**	**0.282 ± 0.034**
	Combo	0.25	**0.980 ± 0.004**	**3.902 ± 0.784**	**0.314 ± 0.054**	0.980 ± 0.003	3.893 ± 0.545	0.304 ± 0.040
		0.5	0.980 ± 0.004	3.979 ± 0.685	0.320 ± 0.045	0.981 ± 0.003	3.788 ± 0.512	0.296 ± 0.041
		0.75	0.978 ± 0.004	4.277 ± 0.702	0.345 ± 0.049	0.978 ± 0.003	4.221 ± 0.511	0.331 ± 0.038

Table 5.3: Results of Pearson correlation (R) and Bland-Altman analysis. Vertex-wise thickness and T2 measurements from manual segmentations were compared to the respective vertex-wise measurements from automated segmentations of both networks. *Mean bias and 95% limits of agreement (LoA) are in mm. ** Mean bias and 95% LoA are in ms

Parameter	Network	Cartilage Surface	R-Value	Mean Bias [95% LoA]
Thickness*	2D U-Net	Femoral	0.75	0.33 [-0.28; 0.96]
		Lateral Tibial	0.89	-0.03 [-0.21; 0.15]
		Medial Tibial	0.74	0.12 [-0.16; 0.40]
	3D U-Net	Femoral	0.98	0.07 [-0.11; 0.25]
		Lateral Tibial	0.83	0.11 [-0.14; 0.36]
		Medial Tibial	0.85	0.14 [-0.10; 0.39]
T2**	2D U-Net	Femoral	0.94	-0.16 [-4.71; 4.40]
		Lateral Tibial	0.97	0.47 [-2.68; 3.62]
		Medial Tibial	0.96	1.32 [-2.19; 4.83]
	3D U-Net	Femoral	0.99	-0.05 [-2.06; 1.95]
		Lateral Tibial	0.98	0.38 [-2.12; 2.87]
		Medial Tibial	0.97	0.46 [-2.47; 3.39]

showed that the mean bias [95% limits of agreement] for femoral cartilage thickness measurements was 0.33 [-0.28; 0.96] with 2D U-Net, however only 0.07 [-0.11; 0.25] mm with 3D U-Net. The 2D U-Net lateral tibial thickness measurement demonstrated a very small underestimation while those from 3D U-Net a slightly larger overestimation (Table 5.3, Figure 5.4). Both 2D and 3D U-Net demonstrated a similarly small systematic overestimation of medial tibial cartilage thickness. The mean biases and 95% limits of agreement for femoral and tibial T_2 measurements showed similar trends between 2D and 3D U-Net but were generally lower for 3D U-Net (Table 5.3, Figure 5.7).

Vertex-wise root-mean-squared (RMS) thickness measurements extracted from all cartilage surfaces generated from 2D U-Net segmentations of the OAI ZIB testing set demonstrated moderate-to-high linear correlations (range: R = 0.74 - 0.89, Figure 5.3), while those extracted from 3D U-Net segmentations demonstrated high correlations (range: R = 0.83 - 0.98, Figure 5.3).

Vertex-wise RMS T_2 measurements extracted from all cartilage surfaces generated from 2D U-Net (range: R = 0.94 - 0.97, Figure 5.6) and 3D U-Net (range: R = 0.97 - 0.99, Figure 5.6) segmentations demonstrated high linear correlations with the measurements obtained from manual segmentations.

Surface-averaged RMS thickness and T_2 measurements of all cartilage surfaces extracted from manual and automatic segmentations of the OAI ZIB testing set using 3D-CaSM are listed in Table 5.4. Vertex-wise RMS thickness measurements and T_2 relaxation times extracted from manual and automated segmentations displayed on the canonical femoral, medial tibial and lateral tibial cartilage surfaces are shown in Figure 5.5 and Figure 5.8, respectively. Difference (Δ) maps between data extracted from manual and 2D and 3D U-Net automated segmentations are also shown.

Table 5.4: Results of thickness and T2 analysis of the baseline OAI ZIB testing dataset. Cartilage surface means (standard deviations) are calculated from all vertex-wise RMS thickness and T2 measurements. *Measurements are in mm. **Measurements are in ms.

Parameter	Cartilage Surface	Manual Surface Mean (SD)	Automated Surface Mean (SD)		Absolute Difference	
			2D U-Net	3D U-Net	2D U-Net	3D U-Net
Thickness*	Femoral	1.97 (0.47)	2.30 (0.41)	2.04 (0.46)	0.33	0.07
	Lateral Tibial	1.71 (0.19)	1.68 (0.20)	1.82 (0.23)	0.03	0.11
	Medial Tibial	1.43 (0.13)	1.54 (0.21)	1.57 (0.21)	0.11	0.14
T2**	Femoral	59.07 (5.95)	58.91 (6.70)	59.02 (6.30)	0.16	0.05
	Lateral Tibial	51.10 (6.55)	51.57 (6.63)	51.48 (6.25)	0.47	0.38
	Medial Tibial	52.63 (6.25)	53.95 (5.44)	53.09 (5.71)	1.33	0.46

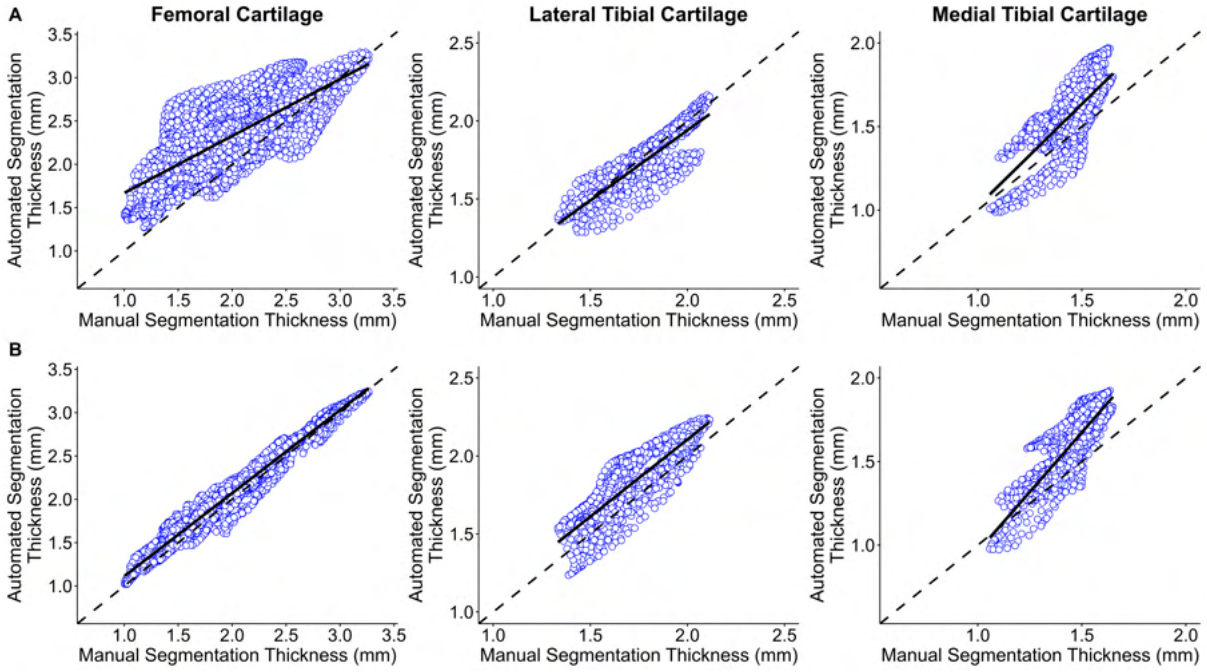

Figure 5.3: Scatterplots show comparison of vertex-wise thickness measurements determined from cartilage surfaces extracted from manual and automatic segmentations (A: 2D U-Net, B: 3D U-Net).

5.4 Discussion

This study aimed to evaluate the use of fully automated segmentations of femoral and tibial bone-cartilage structures for accurate and quantitative cartilage surface-based analysis using the 3D cartilage surface mapping (3D-CaSM) technique.

Compared to an expert human reader requiring 3-4 hours to manually delineate cartilage for a single knee joint, both the 2D and 3D U-Net showed substantially increased time-efficiency by segmenting a complete sagittal 3D DESS volume from the OAI dataset in under 5 s [3]. While both networks showed high agreement with manual segmentation ($DSC > 0.966$), the best performing 3D U-Net ($DSC \geq 0.980$, $ASSD \leq 0.314$ mm) showed a slightly higher accuracy than the best performing 2D network ($DSC \geq 0.971$, $ASSD \leq 0.543$ mm). Therefore, although the training duration for 3D U-Net was substantially larger than that of the 2D U-Nets

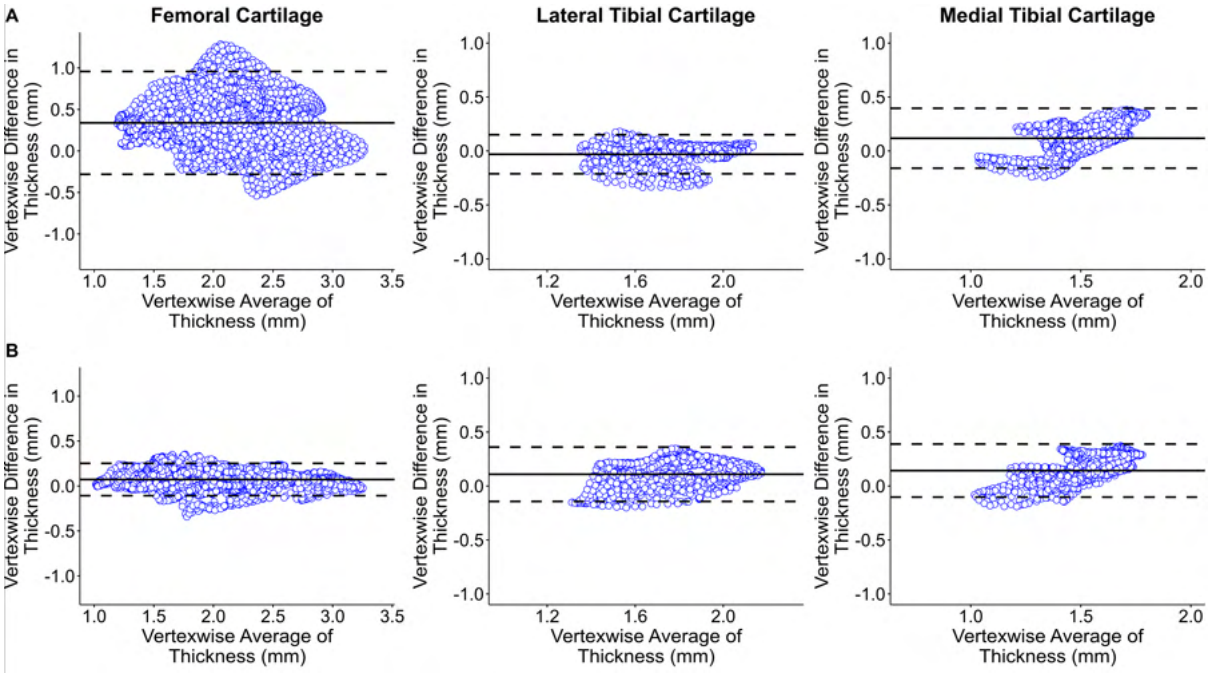

Figure 5.4: Bland-Altman plots show comparison of vertex-wise thickness measurements determined from cartilage surfaces extracted from manual and automatic segmentations (A: 2D U-Net, B: 3D U-Net).

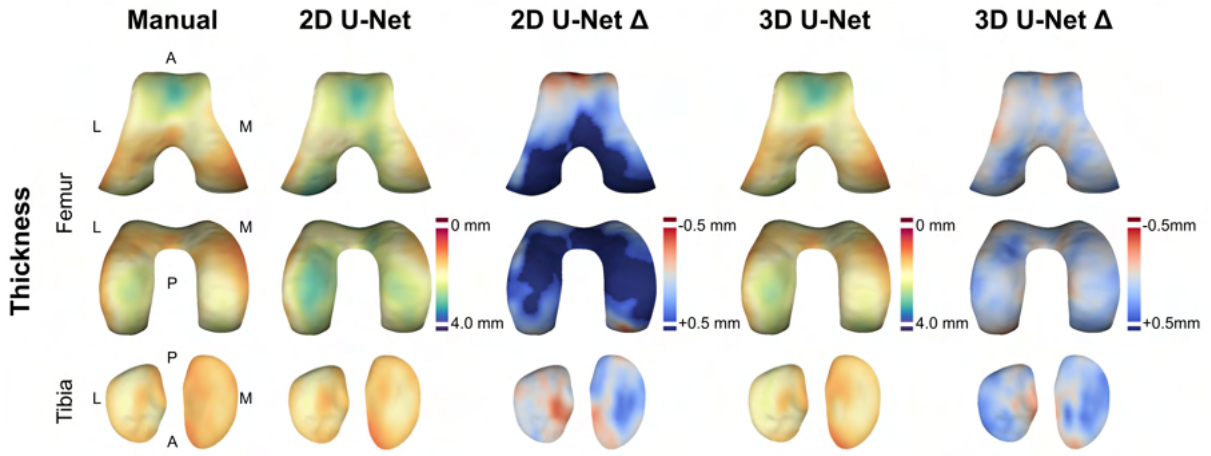

Figure 5.5: Vertex-wise RMS thickness data extracted from manual and automated segmentations from the OAI ZIB test set displayed on the canonical femoral, medial tibial and lateral tibial cartilage surfaces. Difference maps between data extracted from manual and 2D U-Net-automated (2D U-Net Δ) / 3D U-Net-automated (3D U-Net Δ) segmentations highlight spatial regions where the networks experience segmentation difficulties.

due to training on a ten-fold larger volume, applying this fully trained 3D U-Net in practice achieves high performance with only a minor increase in segmentation time. The 3D U-Net used here achieved similar DSCs for both the femoral and tibial bone-cartilage structures compared with those presented by Ambellan *et al* for femoral and tibial bone-only segmentations of the OAI DESS images [153]. However, lower $ASSD$ scores were achieved in this study compared to Ambellan *et al*, which could stem from the more challenging task of segmenting both bone and cartilage in one volume and both tissues having very different contrasts on the DESS images.

We performed 3D cartilage surface mapping (3D-CaSM) to determine cartilage regions where the networks might experience segmentation difficulties compared to manual. When

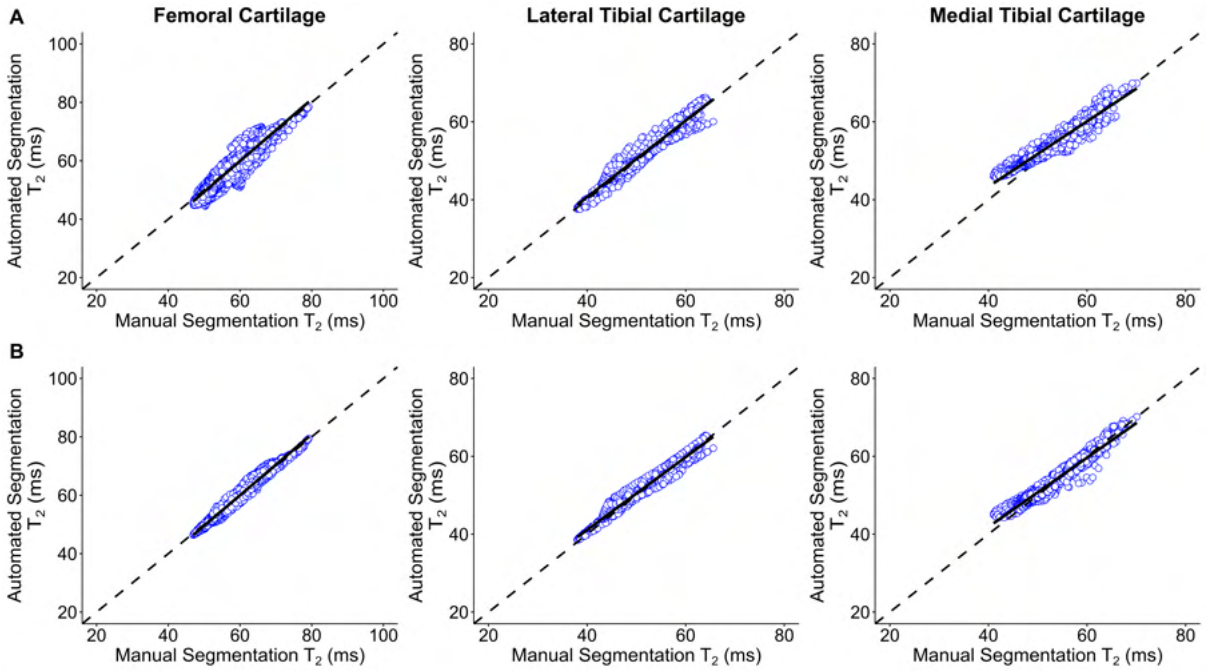

Figure 5.6: Scatterplots show comparison of vertex-wise T_2 measurements determined from cartilage surfaces extracted from manual and automatic segmentations (A: 2D U-Net, B: 3D U-Net).

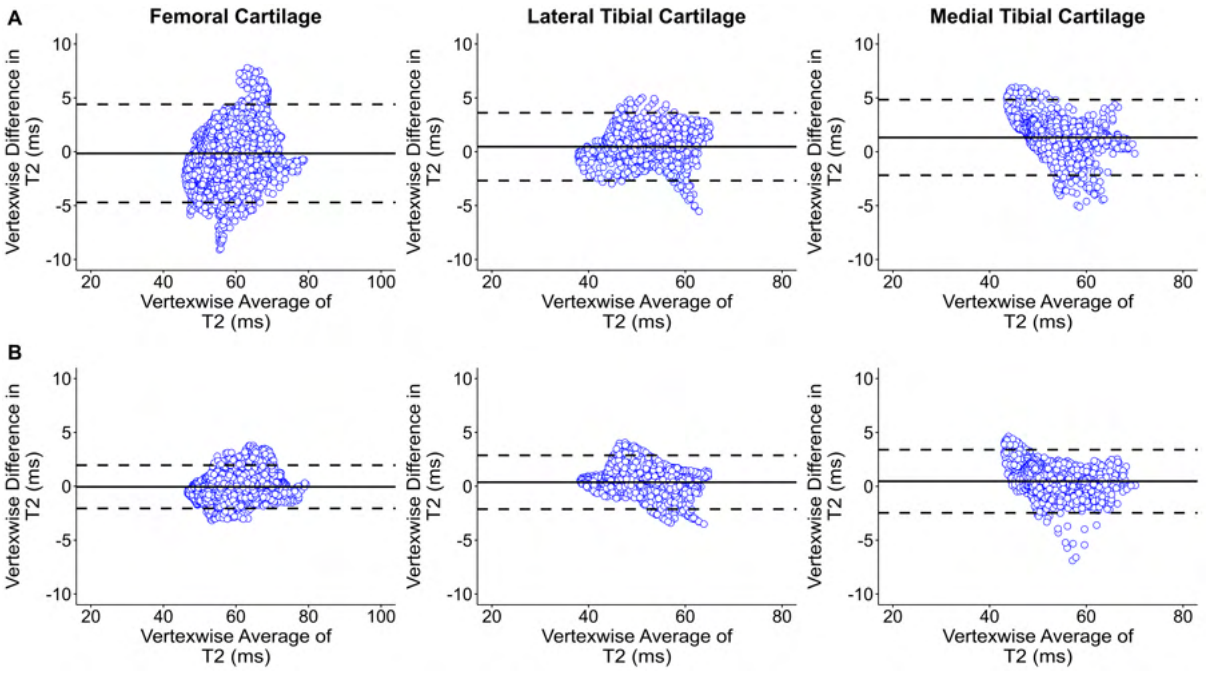

Figure 5.7: Bland-Altman plots show comparison of vertex-wise T_2 measurements determined from cartilage surfaces extracted from manual and automatic segmentations (A: 2D U-Net, B: 3D U-Net).

averaging the thickness measurements across the whole femoral, lateral tibial and medial tibial cartilage surfaces, the determined average femoral cartilage thickness from 2D U-Net segmentations was shown to be substantially larger than the average thickness determined from expert defined manual segmentations (difference = +0.33 mm). By using the 3D-CaSM technique, this average over-estimation of cartilage thickness by the 2D U-Net was regionally localised to the lateral and medial femoral condyle as well as the patellofemoral groove and where these

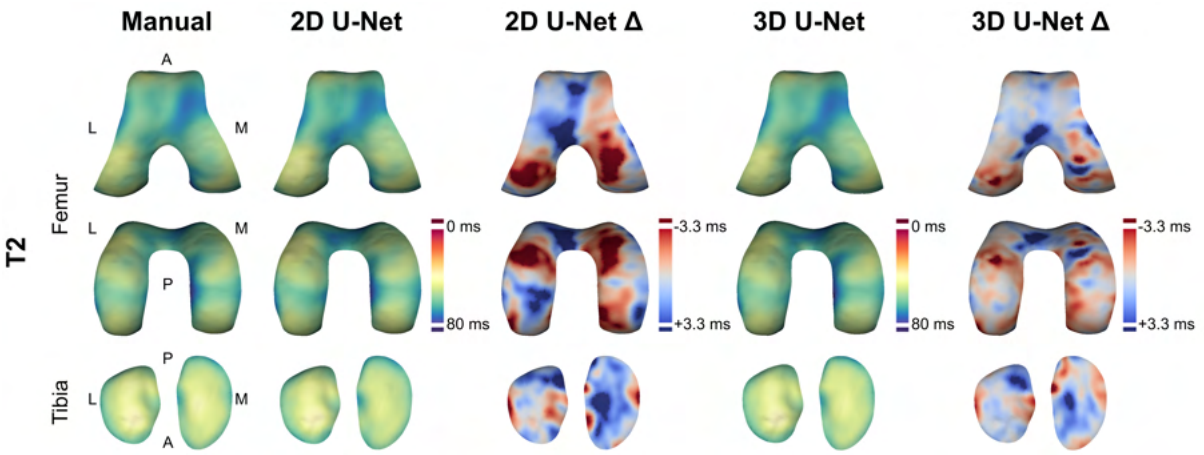

Figure 5.8: Vertex-wise RMS T_2 data extracted from manual and automated segmentations from the OAI ZIB test set displayed on the canonical femoral, medial tibial and lateral tibial cartilage surfaces. Difference maps between data extracted from manual and 2D U-Net-automated (2D U-Net Δ) / 3D U-Net-automated (3D U-Net Δ) segmentations highlight focal regions of T_2 discrepancies.

regions transition into the intercondylar notch. These areas are particularly difficult to segment with different tissues of similar contrast on the DESS images having direct contact to each other such as femorotibial cartilage-cartilage contact and cartilage-meniscus contact. Previous studies have also shown a similar cartilage thickness over-estimation in these regions when using 2D U-Net generated segmentations18,24. However, this regional over-estimation of femoral cartilage thickness was not observed with the 3D U-Net. This could be explained by the 3D network being more volumetrically consistent by training on a larger sub-volume of adjacent image slices instead of individual slices as with 2D networks.

When averaging the T_2 measurements over the entire femoral, lateral tibial and medial tibial cartilage surfaces, differences between manual and 3D U-Net determined T_2 measurements were slightly less than those observed between manual and 2D U-Net. However, when looking at vertex-wise T_2 difference maps between data extracted from manual and automated segmentations, the T_2 2D U-Net difference maps shows higher focal disagreements (T_2 underestimations at the medial and lateral femoral condyles; T_2 over-estimation at the patellofemoral groove) than the T_2 difference maps with 3D U-Net. While the T_2 3D U-Net difference maps of femoral, medial tibial and lateral tibial cartilage show similar patterns to the corresponding T_2 2D U-Net difference maps, the overall vertex-wise T_2 differences with manual segmentations are substantially less pronounced.

While segmentation networks are usually trained on a large amount of imaging data to achieve a sense of generalisability, the networks remain sensitive to small distributional shifts between training and testing data and could produce inaccurate segmentations. Although automated methods lead to more repeatable segmentation results, the proposed method could also be used interactively. Following the automated generation of the structure polygons, they could be manually adjusted by a user to fine-tune and correct the network-generated results. This could not only increase segmentation accuracy, but also substantially reduce the time required by a user to fully delineate structures manually. This could allow the models to be employed by clinical end users and enable the analysis of large medical imaging datasets in an acceptable time.

Although we automated the segmentation of the bone-cartilage structures from which triangulated surface mesh objects were generated, the 3D-CaSM technique used in this work still

required manual outlining of cartilage surface patches. Although this step only required approximately 2 minutes to extract all cartilage patches, future work could focus on fully automating the 3D-CaSM pipeline. Another limitation is the lack of longitudinal validation of 3D-CaSM using contours generated from both manual and network-automated segmentations. This study only assessed the segmentation performance of 2D and 3D U-Net on the OAI baseline dataset as the OAI ZIB dataset only consists of segmentations from baseline OAI scans. Generating manual segmentations from multiple OAI follow-up imaging data could facilitate this limitation in future. Furthermore, the networks were trained, validated and tested on imaging data from the highly controlled OAI observational study. Although the OAI is a multi-centre study, images were acquired on MR systems from a single manufacturer. Additionally, the chosen sample size for testing was relatively small, mainly attributed to limited availability in resources. Future work would aim to expand towards larger and more clinically representational datasets and evaluate clinical significance of imaging biomarkers.

5.5 Conclusion

In the current study we evaluate the use of automated femoral and tibial bone-cartilage segmentations in combination with 3D-CaSM for fast and accurate quantification of cartilage thickness and T_2 relaxation time measurements in comparison with manual segmentation. This method allowed us to highlight localized cartilage regions in which both 2D and 3D U-Nets experienced segmentation difficulties. The presented results demonstrate the validity of using automated segmentations within the 3D-CaSM pipeline to improve the time-efficiency for the MRI-based evaluation of knee joint health.

Chapter 6

Towards Automated Cortical Bone Mapping using Knee MRI and CT Data: Explorative Analysis in the IMI-APPROACH Cohort

6.1 Introduction

Osteoarthritis (OA) is the leading cause of disability in the elderly population and considered a disease encompassing all joint tissues [14]. Understanding the interaction between articular cartilage and the underlying periarticular bone in OA is of fundamental interest in research [190, 191]. Articular cartilage and bone are separated by a layer of calcified cartilage, with the tidemark at the interface. The periarticular bone consists of the cortical bone, the compact subchondral bone closer to the bone surface which transitions into cancellous trabecular bone away from the joint. During the development of OA and disease progression, all these layers undergo structural and compositional changes mainly attributed to alterations in biomechanical effects of loading resulting in bone remodelling [192]. Computed tomography (CT) allows an insight into the OA-induced alterations in cortical and subchondral bone thickness and trabecular bone density triggered by an increased bone remodelling rate in early disease stages. With magnetic resonance imaging (MRI), a direct visualisation of OA-induced cartilage loss, meniscal degeneration, and formation of bone marrow lesions and osteophytes is possible.

Recent research studies evaluating the longitudinal progression of knee OA have incorporated imaging with different modalities to validate OA imaging biomarkers [154, 193]. However, to facilitate the extensive amount of time required to manually analyse these large cohort imaging studies, automated analysis pipelines need to be established. Automating the segmentation process of different parts of the knee joint, especially that of bone and cartilage from MRIs, is an active area of research [153, 164, 183]. The International Workshop on Osteoarthritis Imaging (IWOAI) Knee MRI Segmentation Challenge showed promising results for cartilage segmentations with networks achieving Dice scores of approximately 0.81 – 0.90, with all teams using deep neural network–based models [194]. To achieve a segmentation model with high generalisability through supervised learning, typically a large dataset containing a diverse range of disease representations and high-quality segmentation maps are required. However, when only a small amount of labelled data is present, transfer learning has been introduced as

a means of still achieving a sense of generalisability on that dataset by fine-tuning a pretraining model on a larger dataset with a related task [144, 195].

Surface-based approaches have recently been shown to be particularly efficient in their ability to accurately measure cortical bone thickness from hip CTs as well as joint space width from knee and ankle CTs [196, 197]. A similar approach was also used for measuring cartilage thickness from knee MRIs [3]. A 3D CT analysis technique termed cortical bone mapping has previously been shown to provide accurate and reliable quantitative measures of cortical bone thickness of the hip bone at different OA disease stages [198, 199]. This method could therefore also be applied to the morphological evaluation of the periarticular structures of the knee joint and allow a characterisation of the structural changes occurring at the cartilage-bone interface during OA disease progression. However, the periarticular bone surface cannot be directly delineated from CTs of the knee joint. Nevertheless, it is well definable on MRIs due to a greater soft tissue contrast of articular cartilage.

Using these recently developed strategies, the aim of this study was to assess the feasibility of using an automated pipeline for subchondral cortical bone thickness measurements from CTs of the knee joint. The pipeline is based on automated bone and cartilage segmentations from knee MRIs using a deep neural network from which a 3D periarticular bone surface is extracted and used for cortical bone mapping of corresponding CTs.

6.2 Methods

6.2.1 Image Datasets

OAI-ZIB

This dataset comprises of 507 patients from the publicly accessible Osteoarthritis Initiative (OAI) cohort [153, 154]. The OAI was approved by the Committee on Human Research, the Institutional Review Board for the University of California, San Francisco (UCSF) with all participants providing written informed consent. The OAI-ZIB cohort involved participants with all degrees of OA progression (Kellgren-Lawrence (KL) 0 – 4), with a larger shift towards participants with moderate-to-severe OA (KL $\geq$ 3). Detailed characteristics of the OAI-ZIB cohort are in Ambellan et al [153]. MRIs from the OAI cohort were acquired on 3.0 T Siemens Magnetom Trio systems using a sagittal 3D double echo in steady state (DESS) sequence with water excitation. Sequence parameters were: FOV = 140 mm^2, reconstructed matrix size = 384 x 384, reconstructed in-plane resolution = 0.36 x 0.36 mm^2, slice thickness = 0.7 mm, repetition time = 16.3 ms, echo time = 4.7 ms, flip angle = 25°.

APPROACH

The Applied Public-Private Research enabling OsteoArthritis Clinical Headway (APPROACH) study is a multi-centre observational cohort study with 2-year follow-up clinical, biomechanical, biochemical and imaging data of 297 knee OA patients [193]. Multiple follow-up MRI and CT data were acquired across five different European sites. MRIs were acquired on 3.0 T MR systems from different vendors (Siemens and Philips) using sagittal 3D spoiled gradient echo (3D GRE) based sequences with closely-matching acquisition parameters at the various sites. Equally, CTs were acquired on CT systems from different vendors (Siemens, Philips and

GE) with comparable acquisition parameters between sites. Details on MRI and CT acquisition parameters are in Tables 1 and 2, respectively.

Table 6.1: MRI pulse sequence parameters of training, validation and testing datasets.

	Paris	Utrecht	Leiden	Oslo	A Coruña
MR System	Siemens Skyra	Philips Achieva	Philips Ingenia	Siemens Aera	Philips Ingenia CX
3D Spoiled GE Sequence	FLASH	FFE	FFE	VIBE	FFE
Repetition Time (ms)	17	17	17	17	17
Echo Time (ms)	7	7	7	7	7
Flip Angle (°)	12	12	12	15	15
Field-of-View (mm_2)	160	160	150	160	160
Reconstructed Matrix	512 x 512	512 x 512	512 x 512	512 x 512	560 x 560
In-Plane Resolution (mm)	0.31 x 0.31	0.31 x 0.31	0.29 x 0.29	0.31 x 0.31	0.29 x 0.29
Slice Thickness (mm)	1.5	1.5	1.5	1.5	1.5

6.2.2 Data Preparation

OAI-ZIB

Segmentations were performed by specialists at Zuse Institute Berlin (ZIB) by first generating outlines of femoral and tibial bone and cartilage using a statistical shape model, followed by manual adjustments. Only segmentations of femoral bone and cartilage were used in this work, primarily to pre-train the networks used for segmentation of the APPROACH dataset described below. Consequently, the dataset was only split into 487 for network training and 20 for network validation.

APPROACH

The analysis in this work focussed solely on the baseline MRI and CT data. Manual segmentations of femoral bone and cartilage of 35 randomly chosen APPROACH 3D GRE datasets were performed by a single observer. The dataset was split into 15 for training, 5 for validation and

Table 6.2: CT acquisition parameters of testing dataset.

	Paris	Utrecht	Leiden	Oslo	A Coruña
CT System	Siemens Somatom Definition Edge	Philips IQon	Canon; Toshiba Aquilion ONE	Philips Brilliance 16	GE LightSpeed VCT
Reconstruction Diameter	400	400	400	313	400
Reconstructed Matrix	512 x 512	512 x 512	512 x 512	512 x 512	512 x 512
In-Plane Resolution (mm)	0.78 x 0.78	0.78 x 0.78	0.78 x 0.78	0.61 x 0.61	0.78 x 0.78
Slice Thickness (mm)	0.75	0.80	1.0	0.80	0.63
Peak Output (kV)	120	120	120	120	120
X-Ray Tube Current (mA)	23 - 25	17 - 19	24 - 43	97	25
Volume CT Dose Index (mGy)	0.65 - 0.70	1.26 – 1.43	1.50 – 5.00	17.10	-
Convolution Kernel	I40f\3	C	FC09	B	Standard

Table 6.3: Summary of the APPRAOCH dataset characteristics used for training, validation and testing. * Values are means ± standard deviations. Abbreviations: KL, Kellgren–Lawrence grade; P, Paris; U, Utrecht; L, Leiden; O, Oslo; AC, A Coruña

Training	Number of Subjects	15
	Sex [Female / Male]	11 / 4
	Age (years)*	65.7 ± 7.5
	BMI (kg/cm2)*	29.4 ± 5.3
	Subjects per KL Grade [0 / 1 / 2 / 3 / 4]	1 / 1 / 9 / 2 / 1
	Subjects per site [P / U / L / O / AC]	2 / 5 / 4 / 2 / 2
Validation	Number of Subjects	5
	Sex [Female / Male]	5 / 0
	Age (years)*	63.8 ± 5.2
	BMI (kg/cm2)*	30.4 ± 6.5
	Subjects per KL Grade [0 / 1 / 2 / 3 / 4]	1 / 1 / 1 / 1 / 1
	Subjects per site [P / U / L / O / AC]	1 / 1 / 1 / 1 / 1
Testing	Number of Subjects	15
	Sex [Female / Male]	10 / 5
	Age (years)*	66.9 ± 9.9
	BMI (kg/cm2)*	30.9 ± 5.8
	Subjects per KL Grade [0 / 1 / 2 / 3 / 4]	2 / 3 / 3 / 5 / 2
	Subjects per site [P / U / L / O / AC]	2 / 5 / 4 / 2 / 2

15 for testing. The validation set was chosen to include all five KL grades and imaging sites while the training and testing sets were selected to have equal distribution of imaging site (image contrast) however random distribution of KL grades. Semi-automatic segmentations of the femur bone were created from the corresponding CT data of the 15 testing set cases. After generating threshold-based outlines of the femur on the axial CT slices, manual adjustments were performed. Both MRI and CT segmentations were completed using the StradView software v6.1 (https://mi.eng.cam.ac.uk/Main/StradView). Detailed characteristics of the APPROACH training, validation and testing cohorts are in Table 6.3.

6.2.3 Model Architecture and Training

The CNN used here was the same 3D U-Net used in Chapter 5 for segmenting femoral and tibial bone-cartilage structures from knee MRIs.

For network training, each individual dataset was resampled and split into 256 x 256 x 10 sub-volumes of five overlapping slices. This resulted in each OAI 3D DESS dataset being divided into 31 sub-volumes and each APPROACH dataset into 11 – 23 sub-volumes, depending on the number of slices acquired at the different imaging sites. The bone and cartilage segmentations were each stored in a separate channel in the segmentation map.

Networks were trained with a binary cross entropy loss (BCE) and optimised with the Adam

solver with a learning rate of 0.0002 and momentum parameters, β_1=0.5, β_2=0.999. Within PyTorch, the images were converted to tensors and normalized with the sub-volumes mean signal intensity and standard deviation. The networks and optimisation were implemented using PyTorch (Torch v1.0.1) with all network trainings performed on a Nvidia P6000 GPU card (3840 CUDA cores, 24 GB GDDR5X).

As the APPROACH training data only consisted of 15 subjects, we evaluated the use of transfer learning to improve segmentation accuracy on the APPROACH testing data. We assessed the influence of the pre-training / fine-tuning ratio on the segmentation performance by varying the amount of pre-training on the OAI-ZIB dataset between 10 to 90 epochs, followed by network fine-tuning on the APPROACH dataset for 90 to 10 epochs. The results were also compared to the APPROACH testing set evaluated on OAI-ZIB-only and APPROACH-only training with random weight initialisation where early stopping was used as soon as the networks performance did not improve over a period of 20 epochs measured on their respective validation sets.

6.2.4 Postprocessing and 3D Surface Analysis

Periarticular Bone Surface Extraction and Reconstruction

To extract the periarticular bone surface, i.e. the bone surface covered by articular cartilage, we first combined the segmentation masks of bone and cartilage into one structure (Figure 6.1, step 2). Next, contours of the cartilage-only segmentation and the created bone-cartilage segmentation were calculated and subtracted to acquire the mask of the periarticular bone surface (blue line in step 3, Figure 6.1). A point cloud was calculated by transforming the determined periarticular bone mask voxels into the individual patient-based coordinate system corresponding to the DICOM reference coordinate system. Triangulated surfaces were reconstructed from the generated point cloud using the screened Poisson surface reconstruction algorithm implemented in the Open3D package (Figure 6.1, step 4) [200, 201]. The surfaces were reconstructed with a maximum tree depth of 8. Following simple neighbour average smoothing for 5 iterations, the total number of vertices were decreased by downsampling into a voxel size of 1.5 mm (= slice thickness of MRIs). The average number of vertices in a periarticular bone patch reconstructed from the APPROACH testing cohort was 5500 (range: 4800 - 6700) depending on the extent of individual periarticular bone surfaces. This surface extraction and reconstruction pipeline was processed within Python (v3.7.6, Python Software Foundation, Wilmington, Delaware, United States).

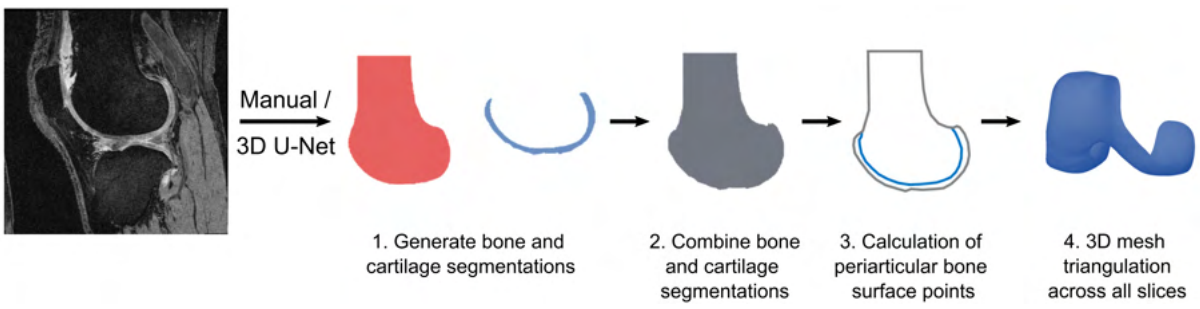

Figure 6.1: Flowchart of periarticular bone surface estimation. Manual or 3D U-Net automated bone and cartilage segmentations are combined to form a single bone-cartilage structure. The periarticular bone surface points are then calculated by subtracting the contours from cartilage-only and bone-cartilage structures. A 3D triangulated periarticular surface patch is then generated from a point cloud determined from all periarticular mask voxels.

Registration to CT Bone Surface

Triangulated surface meshes for the whole femur bone were created within StradView from the threshold-based semi-automated segmentations of femoral bone from CT data as well as manual and automated segmentations of femoral bone from MRI data. In the case of the automated MRI bone segmentation, contours from 3D U-Net generated masks of femoral bone structures were extracted and converted into tissue-specific polygons within Python and stored in a StradView readable file prior to surface reconstruction. For each individual in the APPROACH testing cohort, the femoral bone MR surfaces were registered to the corresponding femoral bone CT surface using a similarity transformation with an iterative closest point (ICP) registration algorithm. Surface registrations were completed using the freely available wxRegSurf software v20 (http://mi.eng.cam.ac.uk/ ahg/wxRegSurf). The resulting similarity transformation matrix for registering whole bone surface was subsequently applied to the corresponding periarticular bone patch determined from the MRI data. The registered periarticular bone surfaces were then used to perform subchondral cortical bone thickness measurements from CT data. The described pipeline is illustrated in Figure 6.2.

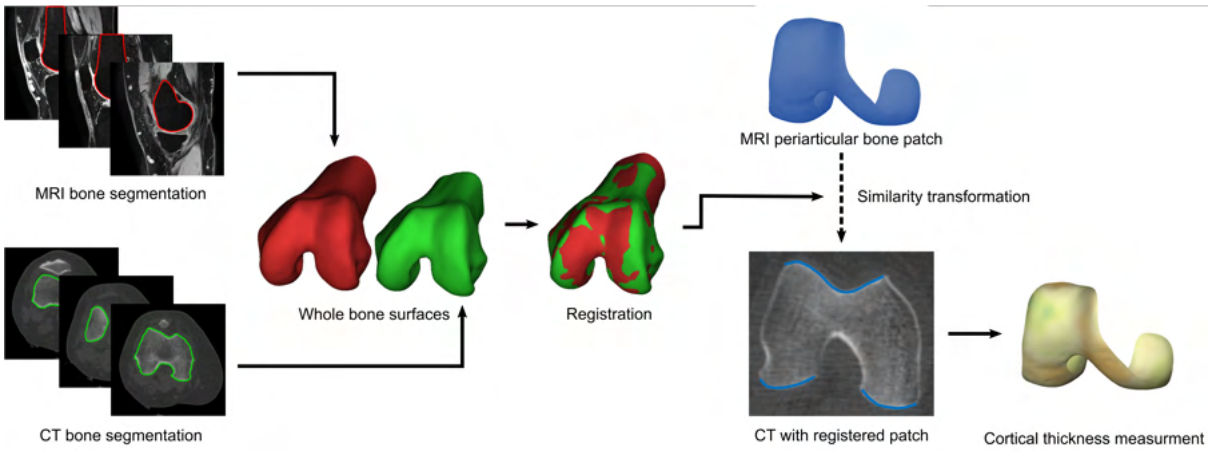

Figure 6.2: Flowchart of cortical bone mapping pipeline. Following the acquisition of MR and CT imaging data, bone is segmented manually or automatically from MRIs and threshold-based semi-automatically from CTs. Whole bone surfaces are generated from the MRI and CT bone segmentations and registered to each other. Using the similarity transformation matrix from the bone registrations, the periarticular bone surface patch is brought into the correct CT space. The last step consisted of automatic cortical thickness measurement at each surface vertex.

Cortical Bone Mapping and Analysis

The next step in the workflow is to perform cortical bone mapping from the periarticular bone surfaces created from manual and automated segmentations for each individual. Cortical bone mapping uses clinical CT imaging data to create 3D surface maps of subchondral cortical thickness. Cortical thickness is measured at each vertex in the triangulated periarticular bone surface patch by sampling the interpolated CT data along a line in the image plane normal to the surface. A normalised Gaussian function is then fitted to the sampled CT data using a Levenberg-Marquardt optimisation method [121, 202].

For each participant, the unique femoral periarticular bone surface mesh was registered to an average (canonical) femoral surface to achieve vertexwise correspondence across individuals. This canonical surface contains roughly 3700 vertices and was calculated from extracted

femoral cartilage surfaces of ten young, healthy participants in a study assessing surface-based changes in articular cartilage composition following a mild exercise regime [186]. Registration was performed by first using a similarity transform (ICP), followed by a thin plate spline transformation in which the rims of the two surface objects are matched [203, 204]. Individual vertexwise cortical thickness measurements were then mapped onto the average surface at the respective closest neighbouring vertices. Both surface registration and mapping of thickness measurements was performed using wxRegSurf software. Vertexwise root-mean-squared (RMS) thickness was calculated from the mapped individual cortical thickness measurements of the 15 APPROACH testing cases.

6.2.5 Model Evaluation and Statistical Analysis

Performance evaluation was assessed in network segmentation accuracy and subchondral cortical thickness measures. For accuracy assessment between manual and 3D U-Net generated segmentations, the Sørensen–Dice Similarity Coefficient (DSC) as well as the boundary distance-based metric Average Symmetric Surfaces Distance (ASSD) were computed [89, 161].

Surfacewise and vertexwise scatterplot and Bland-Altman analysis were performed to evaluate any related differences between cortical thickness measurements from manual and automated segmentations. For surface-wise analysis, a surface-averaged thickness measurement is used, while for vertexwise analyses, the RMS cortical thickness measurements at each canonical surface vertex from the periarticular bone surfaces extracted from manual segmentations were compared to the matching vertexwise measurements on the periarticular bone surfaces determined from automated segmentations.

6.3 Results

6.3.1 Segmentation Performance Comparison

Bone and cartilage segmentation results (DSC and ASSD) of networks trained without and with different intervals of transfer learning on the OAI ZIB are shown in page 110 and Figure 6.4, respectively. The best performing network on both bone and cartilage was the 3D U-Net pre-trained on OAI-ZIB for 40 epochs and fine-tuned for 60 epochs on the APPROACH dataset. The mean DSC of bone and cartilage segmentations between manual ground truth and 3D U-Net was 0.978 ± 0.002 and 0.833 ± 0.033. The mean ASSD of bone and cartilage segmentations between manual and 3D U-Net was 0.325 ± 0.047 mm and 0.324 ± 0.067 mm. For both bone and cartilage, the ASSD was comparable to the in-plane resolution of the training images.

6.3.2 Cortical Bone Mapping

Surfacewise cortical thickness measurements using periarticular bone surfaces were similar between those determined from manual and 3D U-Net automated segmentations. Bland-Altman analysis (Figure 6.5) demonstrated a small positive mean bias [95% limits of agreement] of 0.004 [-0.110; 0.119] mm indicating a minor systematic overestimation of surface-averaged cortical thickness using the automated segmentation approach. The surface-averaged RMS cortical thickness determined with periarticular bone surfaces generated with manual segmentations was 2.04 ± 0.29 mm and 2.04 ± 0.27 mm with 3D U-Net segmentations.

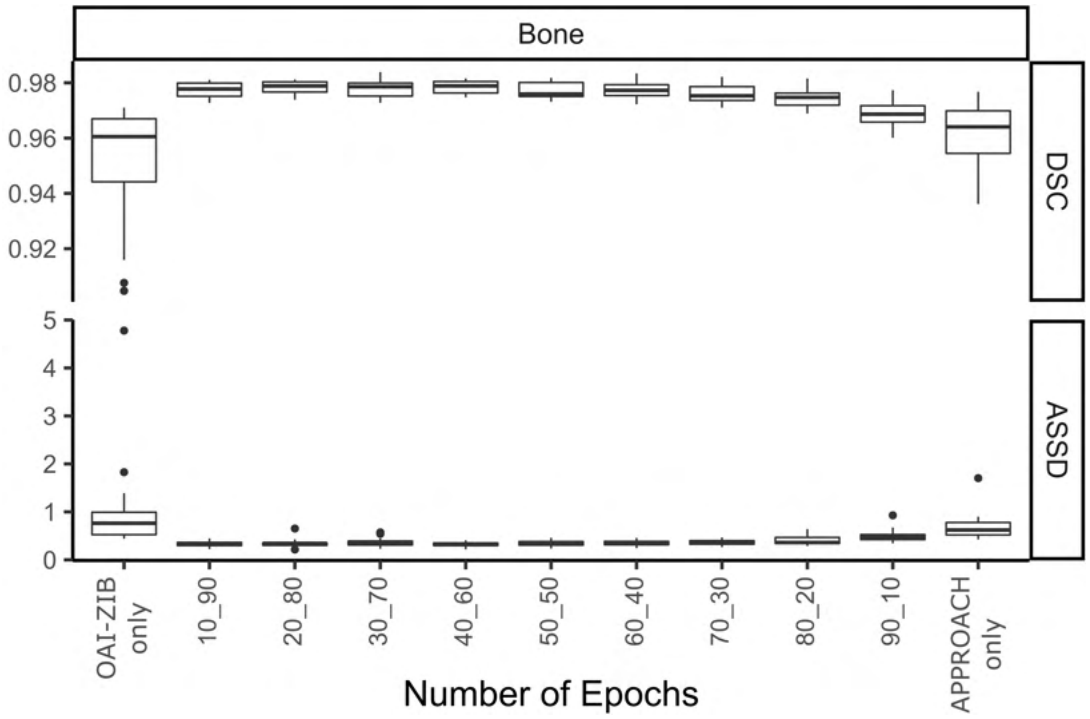

Figure 6.3: Bone segmentation accuracy of the APPROACH testing dataset with different pretraining / fine-tuning ratios. Accuracies were measured with the Sørensen–Dice similarity coefficient (DSC) and average symmetric surface distance (ASSD).

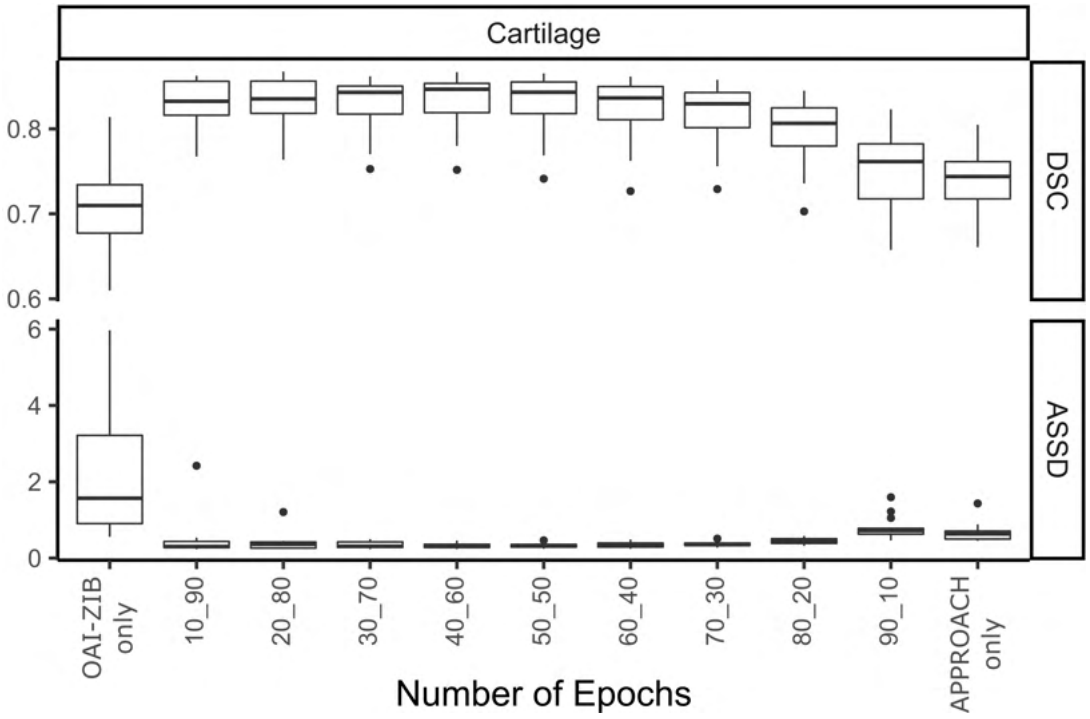

Figure 6.4: Cartilage segmentation accuracy of the APPROACH testing dataset with different pretraining / fine-tuning ratios. Accuracies were measured with the Sørensen–Dice similarity coefficient (DSC) and average symmetric surface distance (ASSD).

Vertexwise cortical thickness measurements extracted from the manual and automated segmentation approaches displayed on a canonical femoral surface are shown in Figure 6.6. Cortical thickness difference maps are also shown. Scatter- and Bland-Altman-plots in Figure 6.7

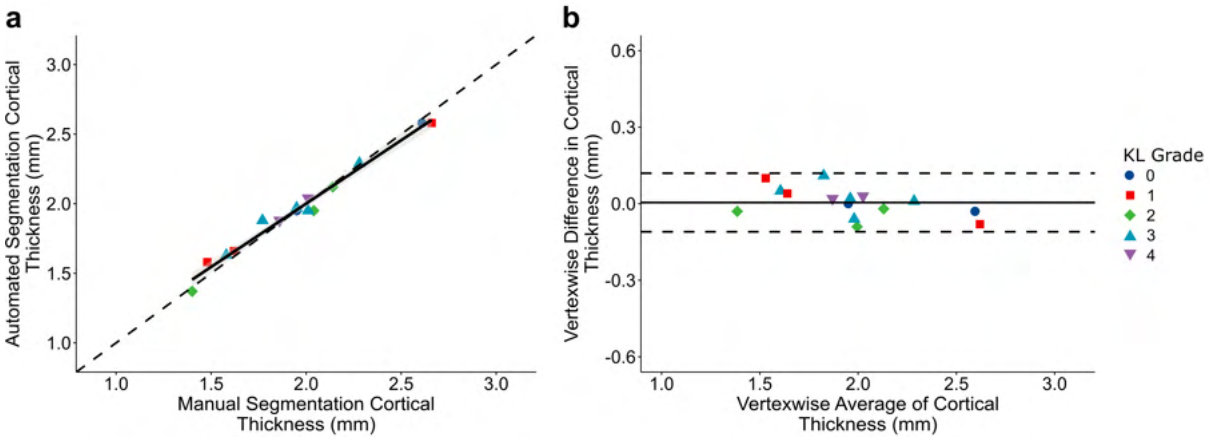

Figure 6.5: Scatterplot and Bland-Altman plot showing the comparison of surfacewise average cortical thickness measurements determined from periarticular bone surfaces extracted from manual and automatic segmentations.

compare vertexwise RMS cortical thickness measurements from manual and automated segmentations. These demonstrate a minor underestimation of femoral subchondral cortical thickness measured with the periarticular bone surfaces determined from 3D U-Net segmentations with a small negative mean bias [95% limits of agreement] of -0.002 [-0.210; 0.205] mm.

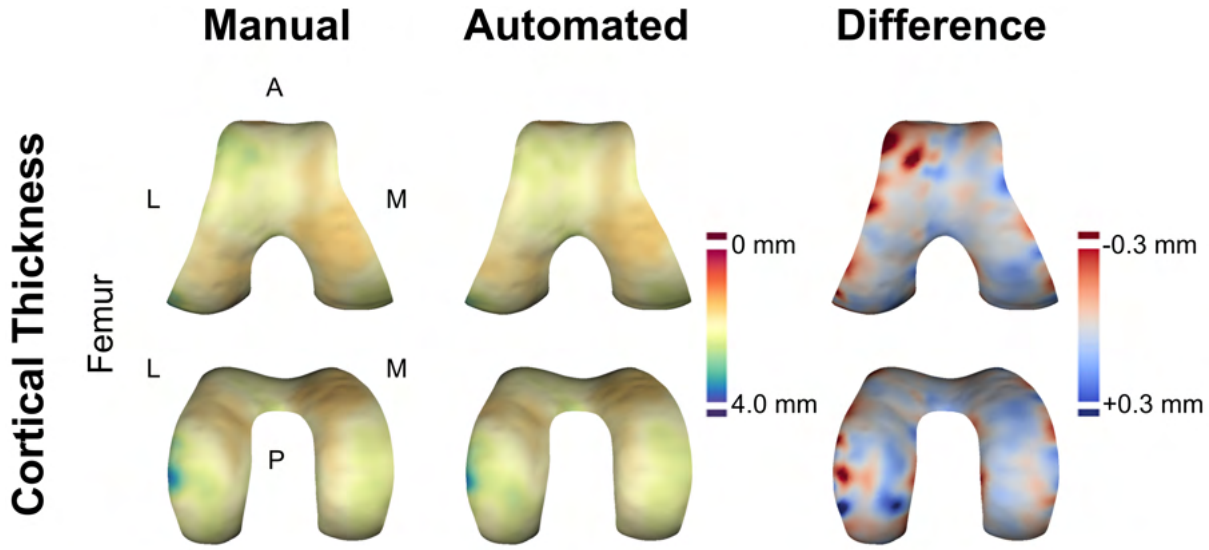

Figure 6.6: Vertexwise RMS cortical thickness data determined from periarticular bone surfaces extracted from manual and 3D U-Net automated segmentations from the APPROACH test set displayed on the canonical femoral surface. Difference maps highlight spatial discrepancies between the manual and automated approaches.

6.4 Discussion

This feasibility study presented a deep neural network–based pipeline for automated bone and cartilage segmentation from knee MRIs which were used to determine a periarticular bone surface applied to cortical bone thickness mapping on corresponding CTs.

This study used 3.0 T MRI data from five MR systems at five different imaging sites to train and test a 3D U-Net. Although we only used 20 APPROACH datasets for training / validation, the performance of bone and cartilage segmentations using a 3D U-Net demonstrated high

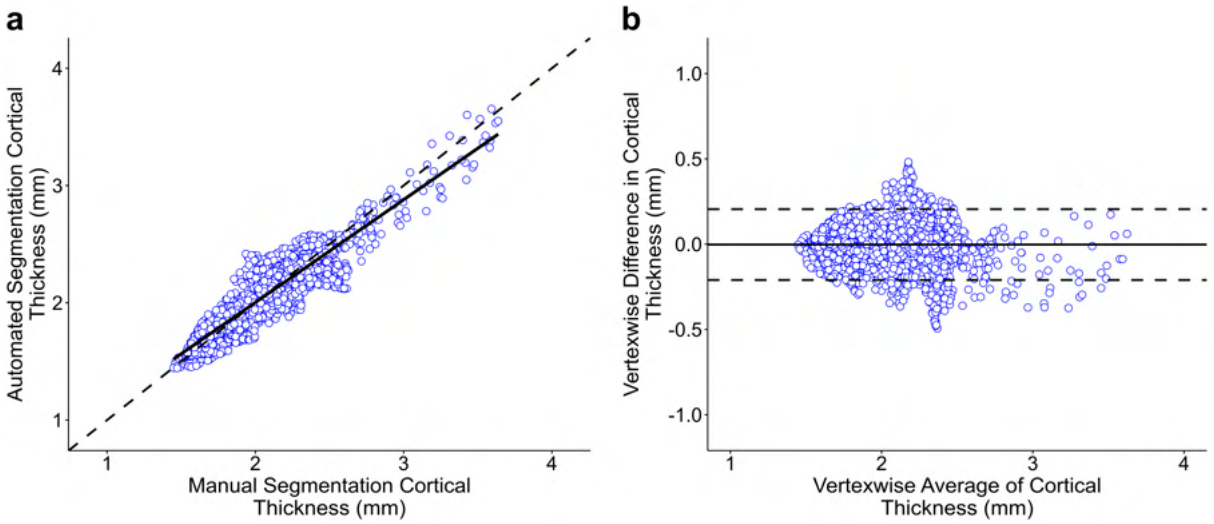

Figure 6.7: Scatterplot and Bland-Altman plot showing the comparison of vertexwise cortical thickness measurements determined from periarticular bone surfaces extracted from manual and automatic segmentations.

accuracy with manually performed segmentations. We showed that with network pretraining on the much larger OAI ZIB dataset, segmentation performance on the APPROACH dataset was considerably improved. We used a binary cross entropy loss function in this study which resulted in high segmentation performance. Although there are many other loss function alternatives used for segmentation purposes, a complete analysis to determine the best loss function for our task was beyond the scope of this feasibility study.

Surfacewise and vertexwise cortical bone thickness measurements using periarticular bone surfaces determined from automated segmentations were similar to those determined from manual segmentations indicated by the small biases from Bland-Altman analysis. Additionally, the 95% limits of agreement were substantially lower than the voxel resolution of the CT data for both surfacewise and vertexwise analysis. A strength of this study is the use of two heterogeneous datasets, the OAI ZIB and APPROACH, for network training, with both consisting of MRI scans of variable image quality and contrast. This could potentially allow the networks to achieve high segmentation quality in a clinical setting. This would however require a more comprehensive validation. A next step could be to apply this method to analyse the entire APPROACH dataset and evaluate its longitudinal and clinical efficacy. This would allow an evaluation of the relationship between femoral periarticular cortical bone thickness and radiological knee OA, similar to what has previously been performed at the hip [199].

The proposed method could additionally be used to characterise the effects of different OA management and treatment approaches on cortical bone remodelling. While a clear beneficial response of exercise programs on cartilage restoration in patients with OA has not yet been determined, one study has shown an improvement of cartilage composition in an asymptomatic, untrained female cohort following a 10-week running program using quantitative MRI [205]. A recent study determined a significant increase in femorotibial cartilage after two years following intra-articular administration of a disease modifying OA drug every six to twelve months [206]. A more invasive approach known as knee joint distraction has shown significant clinical improvements and cartilage regeneration in an end-stage knee OA cohort up to ten years following treatment [207]. Using the proposed method to additionally study the remodelling response of the periarticular cortical bone to various therapeutic approaches in individuals with knee OA could assist in understanding their potential benefits on the cartilage-bone unit as

a whole. It would allow insight into the temporal and spatial relationship between cartilage damage and changes in underlying bone. Such analysis would however require CT and MRI data at spatially corresponding resolutions and ideally at more granular scales to allow greater precision than previously achievable.

We recognise that this study has limitations. The feasibility of our cortical thickness measuring pipeline was only assessed at the femoral periarticular bone surface. An evaluation of using the proposed automated pipeline for tibial and patellar periarticular cortical bone mapping would be the next step to enable a full knee joint assessment. A further limitation of this study was that only a single observer was used for manual data segmentation. We did not assess intra- or interobserver reliability of this method at this point as the aim was to make the entire pipeline operator independent.

6.5 Conclusion

This work demonstrates the feasibility of an automated analysis pipeline for accurate subchondral cortical bone thickness measurement at the femoral periarticular bone surface. Deep learning enabled automated bone and cartilage segmentations which allowed the estimation of the periarticular bone surface from knee MRIs. Subsequently, this surface could be applied to corresponding knee CTs for accurate cortical thickness measurement making the analysis process of large clinical datasets such as APPROACH feasible, operator independent and faster by reducing manual burden.

Chapter 7

Future Developments and Conclusions

7.1 Contributions to Knowledge

This thesis has investigated the role of quantitative magnetic resonance imaging (MRI) and medical image analysis techniques for evaluating the knee joint tissues with a focus on cartilage and osteoarthritis. The primary contributions of this thesis include the detection of changes in cartilage composition following exercise-induced joint loading, an evaluation of the use of convolutional neural networks (CNN) for segmenting multiple knee joint tissues simultaneous from MRIs, and assessments of the performance of CNNs for automated and time-efficient analysis of cartilage morphology and composition as well as cortical bone thickness.

Measuring the compositional response of the femoral, tibial, and patellar articular cartilage after a mild stepping activity and comparing these changes to the intrasessional repeatability of $T_{1\rho}$ and T_2 relaxation time mapping allowed the detection and localisation of focal changes in cartilage composition greater than measurement error (Chapter 3). By measuring the intrasessional repeatability of $T_{1\rho}$ and T_2 relaxation mapping, this work has provided the smallest detectable differences in $T_{1\rho}$ and T_2 relaxation time for use in future studies. Imaging the compositional recovery after exercise has shown slightly greater recovery of $T_{1\rho}$ compared to T_2, demonstrating a possibly faster recovery of the proteoglycan content compared to the collagen organisation within cartilage. This work provides an important step towards understanding the immediate effects on and recovery of articular cartilage following an acute joint loading activity and could be informative for establishing appropriate exercise programs for the management of OA.

The optimisation and validation of various CNNs for segmenting several knee joint tissues from MRIs in Chapter 4 has provided technical information required to improve our understanding of the functionality and implementation of these networks. This work has shown that, with meticulous optimisation, conditional generative adversarial networks (cGANs) can achieve equivalent segmentation performance to the well-known U-Net technique. This is an important step for future developments of cGANs and U-Nets and lays a foundation for utilising these networks for segmentation purposes.

The development and assessment of the automated segmentation pipeline presented in Chapter 5 highlighted the importance of understanding in which knee joint regions CNNs perform well and in which regions they experience segmentation difficulties. Knowledge of the limitations can avoid misinterpreting under- or overestimation of cartilage thickness by a CNN as cartilage de- or regeneration, respectively. The major benefit of the developed analysis

pipeline lies within the greatly improved time-efficiency for analysing cartilage morphology and composition allowing it to next be applied to the Applied Private-Public partneRship enabling OsteoArthritis Clinical Headway (APPROACH) cohort and potentially establish clinically meaningful imaging outcomes.

The automated pipeline for cortical bone thickness measurement developed in Chapter 6 has set an initial step in making the analysis of large-scale longitudinal OA imaging datasets feasible. This analysis pipeline could in future help towards improving our understanding of the associated changes in the bone-cartilage unit during OA progression. Although the network was trained, validated and tested on a limited number of subjects in the APPROACH dataset, high agreement was achieved between the cortical thickness measurements determined from the periarticular bone patch extracted from manual and automated segmentations, indicating a robust CNN and pipeline.

7.2 Future Developments

The work presented in Chapter 3 was part of the protocol and analysis development for a prospective exploratory OA imaging study termed Functional Exercise Response on osteoArthritis Relaxation Imaging (FERARI). The main purpose of the study is to develop and determine a sensitive MR imaging biomarker that differentiates between participants with healthy knee cartilage and patients with early osteoarthritis (OA). Additionally, the magnitude of changes in compositional MRI values will be determined over 1 year, in the absence of any disease-modifying intervention. The study consists of imaging each participant prior to exercise, during mechanical loading of the knee joint using an MR compatible loading device (Diagnostic Pedal from Ergospect GmbH), and after the stepping exercise proposed in Chapter 3. Similarly, after exercise, multiple time points will be collected in the following hour to measure joint recovery Figure 7.1.

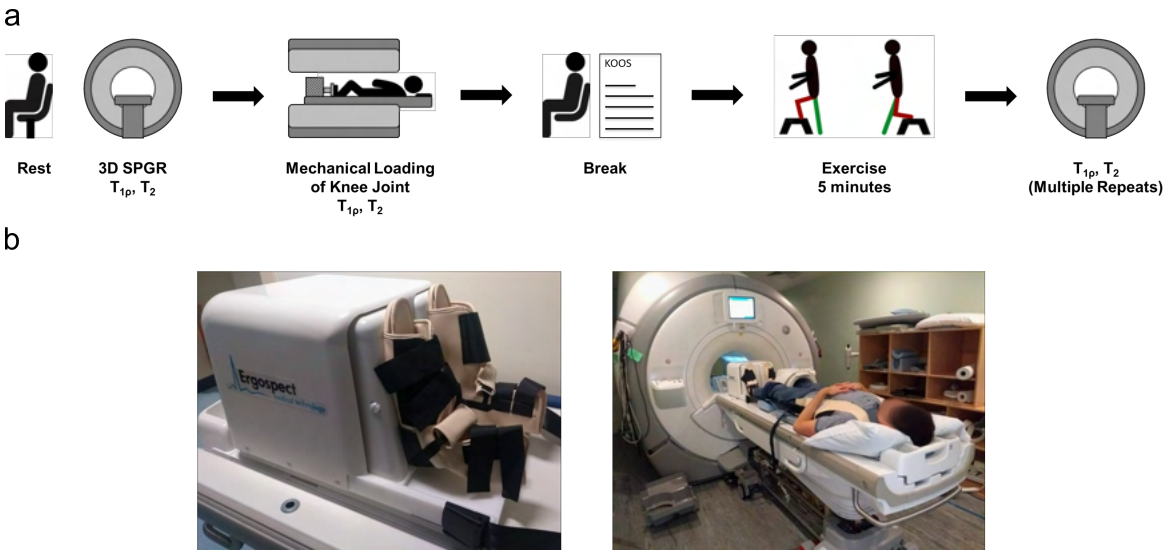

Figure 7.1: FERARI study protocol (a): After an initial rest period, clinical and pre-loading research imaging is performed. While still in the MRI scanner, the knee joints of the participant are mechanically loaded with an MR compatible loading device (b) for approximately 10 minutes followed by post-loading research imaging. After a 30-minute rest period to mitigate the effects of the mechanical loading on the knee joint, participants perform a 5-minute stepping activity followed by post-exercise recovery imaging.

Additional information, such as gender, age, weight, height, and weekly exercise activity, will be collected to observe correlations beyond OA severity. Participants (12 OA, 6 normal volunteers) were selected according to the following main inclusion criteria: (1) aged 40 - 65 years; (2) body mass index $\leq$35 kg/m^2; (3) imaged knee able to fit within the knee MRI coil (approximately 18 cm diameter); and (4) able to perform exercise component of the examination. Additional inclusion criteria for OA participants were a clinical diagnosis of OA per American College of Rheumatology criteria and mild medial tibiofemoral predominant disease with a Kellgren–Lawrence (KL) system grade 2 assessed on an existing knee radiograph acquired within the previous 12 months. Normal volunteers are matched (1:2 ratio) to OA participants for age, sex and body mass index who do not have knee pain. This study will provide test-retest reliability measurements of quantitative MRI methods promising for assessing early degenerative changes in cartilage composition associated with osteoarthritis as determined from a 1-month repeatability visit. Additionally, the study will provide knowledge of effect sizes for sample size calculations for the use of imaging endpoints in future longitudinal or interventional studies using MR imaging outcome measures.

The quantitative MR relaxation time mapping techniques studied in this thesis can provide more repeatable results with a closer insight into the relationship between the underlying biology and the related MRI-derived values than conventional contrast-weighted imaging. Rather than acquiring a single MR image with only contrast information, quantitative MRI techniques aim at determining quantitative tissue-based MR characteristics. Mapping of quantitative MR relaxation times is promising for improving disease diagnosis at early stages, monitoring progression, and assessing treatment response beyond simple qualitative assessments [23]. However, quantitative imaging is often slow and inefficient as it involves the serial acquisitions of multiple contrast-weighted images. A mathematical model is then fitted to these time-series images from which a single quantitative map can be derived to provide information of a single parameter. Sequential measurement of multiple MR parameters is almost always time-consuming and prone to movement artefacts. MR Fingerprinting (MRF) was recently introduced, as a novel acquisition and reconstruction strategy to overcome these challenges and time-constraints of quantitative imaging techniques [208]. MRF has the potential to be used for clinical imaging as it enables fast, simultaneous, and efficient multi-parametric mapping by exploiting the transient signals produced from the pseudo-random variation of sequence parameters such as repetition time (TR) and flip angle (FA). These generated signal evolutions or 'fingerprints' are unique for different tissues and are dependent on the various MR properties of the tissue, such as proton density or T_1 and T_2 relaxation. After data acquisition, the signals are matched to a simulated dictionary including (but not limited to) T_1 and T_2 to create quantitative maps. The dictionary is generated based on the MRF sequence set-up used and the simulation of the spin development that may be observed during acquisition. MRF uses a relatively simple pattern recognition algorithm to identify the tissue and its corresponding properties in each voxel. The inner products between the normalised measured signal evolution of each voxel and each normalised dictionary entry are calculated. The dictionary entry returning the maximum value of the inner product is taken as the best representation of the acquired signal evolution. The respective T_1 and T_2 values are consequently assigned to the voxel. The proton density (PD) is calculated as the scaling factor used to match the dictionary simulation with the measured signal evolution. Figure 7.2 shows an example of MRF-acquired PD, T_1 and T_2 maps as well as a generated synthetic SPGR FS images of the knee joint. By incorporating MRF in the FERARI study introduced above, future work has the potential to validate MRF as a fast and reliable alternative to the time-consuming, gold-standard relaxation time mapping methods. The

overall long-term objective is to achieve early detection and characterisation of OA by bringing the quantitative MRI techniques studied in this dissertation into more prevalent clinical use. However, for these methods to make this next step, reliability and multicentre reproducibility need to be demonstrated as well as their potential to impact patient management and follow-on treatment.

The quantitative MR relaxation time mapping techniques studied in this thesis can provide more repeatable results with a closer insight into the relationship between the underlying biology and the related MRI-derived values. Rather than acquiring a single MR image with only contrast information, quantitative MRI techniques aim at determining tissue specific MR characteristics. Mapping of quantitative MR relaxation times is promising for improving disease diagnosis at early stages, monitoring progression, and assessing treatment response beyond simple qualitative assessments [23]. However, quantitative imaging is often slow and inefficient as it involves the serial acquisitions of multiple contrast-weighted images. A mathematical model is then fitted to these time-series images from which a single quantitative map can be derived to provide information of a single parameter. A simultaneous measurement of multiple MR parameters is almost always time-consuming and prone to movement artefacts. MR Fingerprinting (MRF) was recently introduced, as a novel acquisition and reconstruction strategy to overcome these challenges and time-constraints of quantitative imaging techniques [208]. MRF has the potential to be used for clinical imaging as it enables fast, simultaneous and efficient multi-parametric mapping by exploiting the transient signals produced from the pseudo-random variation of sequence parameters such as repetition time (TR) and flip angle (FA). These generated signal evolutions or 'fingerprints' are unique for different tissues and are dependent on the various magnetic resonance properties of the tissue. After data acquisition, the signals are matched to a simulated dictionary including (but not limited to) T_1 and T_2 to create quantitative maps. The dictionary is generated based on the MRF sequence set-up used and the simulation of the spin development that may be observed during acquisition. MRF uses a relatively simple pattern recognition algorithm to identify the tissue and its corresponding properties in each voxel. The inner products between the normalised measured signal evolution of each voxel and each normalised dictionary entry are calculated. The dictionary entry returning the maximum value of the inner product is taken as the best representation of the acquired signal evolution. The respective T_1 and T_2 values are consequently assigned to the voxel. The proton density (PD) is calculated as the scaling factor used to match the dictionary simulation with the measured signal evolution. Figure 7.2 shows an example of MRF-acquired PD, T_1 and T_2 maps as well as a generated synthetic SPGR FS images of the knee joint. By incorporating MRF in the FERARI study introduced above, future work has the potential to validate MRF as a fast and reliable alternative to the time-consuming, gold-standard relaxation time mapping methods. The overall long-term objective is to bring the quantitative MRI techniques studied in this thesis and other potential candidates such as MRF into more prevalent clinical use. However, for these methods to make this next step, reliability and multicentre reproducibility need to be demonstrated as well as their potential to impact patient management and follow-on treatment.

Similarly, deep learning (DL) algorithms, such as the ones studied in this thesis for reducing the time to perform accurate medical image analysis, have not yet achieved widespread implementations in clinical settings. One reason for this is that there is no established regulatory pathway for these algorithms to receive approval for clinical use. Additionally, since training, validation and testing are generally performed on the same dataset at a single institution, the reproducibility and generalisability of the DL-based algorithms are rarely established. Specif-

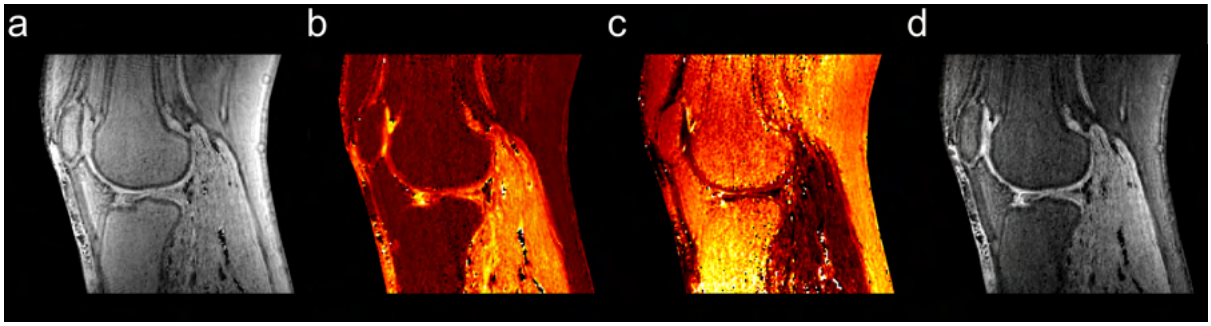

Figure 7.2: Magnetic resonance fingerprinting (MRF) maps of the knee joint from a patient with radiographic knee OA of KL grade 2. a) Proton density, b) T_1, and c) T_2 maps. d) is a synthetic SPGR FS image obtained from the three quantitative MRF maps.

ically to the knee segmentation tasks performed in this thesis, generalisability is difficult to achieve if the datasets trained on are from highly controlled OA imaging studies. Although the Osteoarthritis Initiative is a multi-centre collaboration, all MR imaging was performed on similar MR systems from an individual manufacturer. The APPROACH study incorporates imaging data acquired at five different European sites from various MR systems of two manufacturers allowing an expansion of training deep neural networks on data of more clinical representation. Future work should therefore focus on setting up OA imaging studies with the aim of acquiring international multi-centre, multi-vendor data to achieve not only a more generalisable clinical OA outcome measure but additionally, training of DL algorithms with inherent heterogeneity and widespread applicability. This would allow demonstration of the DL algorithms reproducibility through federated learning across multiple distributed institutions having hold-out local data for testing and potential fine-tune training of the algorithm.

7.3 Conclusions

1. $T_{1\rho}$ and T_2 relaxation time mapping techniques are repeatable in ex-vivo phantoms and in-vivo knee joints.

2. 3D Cartilage Surface Mapping (3D-CaSM) allowed a detection of significant focal changes in cartilage $T_{1\rho}$ and T_2 relaxation times after a uni-lateral stepping activity in a young cohort without knee pain symptoms.

3. With thorough optimisation, conditional generative adversarial networks (cGANs) and U-Nets can achieve comparable segmentation performances for segmenting multiple knee joint structures from various MRI datasets.

4. Both 2D and 3D U-Nets demonstrate the efficiency of using convolutional neural networks (CNNs) for generating accurate segmentations of femoral and tibial bone-cartilage structures. Furthermore, these segmentations can be used to perform surface-based analysis using 3D-CaSM to extract accurate cartilage morphological and compositional features.

5. Using an automated pipeline based on CNN-generated bone and cartilage segmentations from knee MRI data to extract a 3D periarticular bone surface, accurate measurements of cortical bone thickness are achievable on corresponding knee CT data.

Bibliography

1. Cutcliffe, H. C., Davis, K. M., Spritzer, C. E. & DeFrate, L. The Characteristic Recovery Time as a Novel, Noninvasive Metric for Assessing In Vivo Cartilage Mechanical Function. *Annals of Biomedical Engineering* **48,** 2901–2910 (2020).
2. Ebrahimkhani, S. *et al.* A review on segmentation of knee articular cartilage: from conventional methods towards deep learning. *Artificial Intelligence in Medicine* **106,** 101851 (2020).
3. MacKay, J. W. *et al.* Three-Dimensional Surface-Based Analysis of Cartilage MRI Data in Knee Osteoarthritis: Validation and Initial Clinical Application. *Journal of Magnetic Resonance Imaging* **52,** 1139–1151 (2020).
4. Link, T. M. *Cartilage Imaging: Significance, Techniques and New Developments* 258 (2011).
5. Mow, V. C., Lai, W. & Redler, I. Some surface characteristics of articular cartilage—I. A scanning electron microscopy study and a theoretical model for the dynamic interaction of synovial fluid and articular cartilage. *Journal of Biomechanics* **7,** 449–456 (1974).
6. Nordin, M., Frankel, V. H. & Leger, D. *Basic Biomechanics of the Musculoskeletal System* 4th, 472 (2012).
7. Sophia Fox, A. J., Bedi, A. & Rodeo, S. A. The basic science of articular cartilage: Structure, composition, and function. *Sports Health* **1,** 461–468 (2009).
8. Aspberg, A. in *Cartilage: Volume 1: Physiology and Development* (eds Grässel, S. & Aszódi, A.) (Springer International Publishing, 2017).
9. Goldring, M. B., Culley, K. L. & Otero, M. in *Cartilage: Volume 2: Pathophysiology* (eds Grässel, S. & Aszódi, A.) (Springer International Publishing, 2017).
10. Eyre, D. R., Weis, M. A. & Wu, J. J. Articular cartilage collagen: An irreplaceable framework? *European Cells and Materials* **12,** 57–63 (2006).
11. Bertrand, J. & Held, A. in *Cartilage: Volume 2: Pathophysiology* (eds Grässel, S. & Aszódi, A.) (Springer International Publishing, 2017).
12. Bhosale, A. M. & Richardson, J. B. Articular cartilage: structure, injuries and review of management. *British Medical Bulletin* **87,** 77–95 (2008).
13. Sun, H. B. Mechanical loading, cartilage degradation, and arthritis. *Annals of the New York Academy of Sciences* **1211,** 37–50 (2010).
14. Martel-Pelletier, J. *et al.* Osteoarthritis. *Nature Reviews Disease Primers* **2** (2016).
15. Lohmander, L. S., Englund, P. M., Dahl, L. L. & Roos, E. M. The long-term consequence of anterior cruciate ligament and meniscus injuries: Osteoarthritis. *American Journal of Sports Medicine* **35,** 1756–1769 (2007).

16. Simon, L. S. Osteoarthritis: A review. *Clinical Cornerstone* **2,** 26–37 (1999).

17. Sandell, L. J. & Aigner, T. Articular cartilage and changes in arthritis An introduction : Cell biology of osteoarthritis. *Arthritis Research* **3,** 107–113 (2001).

18. Zhu, X., Chan, Y. T., Yung, P. S. H., Tuan, R. S. & Jiang, Y. Subchondral Bone Remodeling: A Therapeutic Target for Osteoarthritis. *Frontiers in Cell and Developmental Biology* **8** (2021).

19. Burr, D. B. & Gallant, M. A. Bone remodelling in osteoarthritis. *Nature Reviews Rheumatology* **8,** 665–673 (2012).

20. Li, G. *et al.* Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. *Arthritis Research & Therapy* **15,** 223 (2013).

21. Buckwalter, J. A. & Martin, J. A. Osteoarthritis. *Advanced Drug Delivery Reviews* **58,** 150–167 (2006).

22. Ismail, H. M. & Vincent, T. L. in *Cartilage: Volume 2: Pathophysiology* (eds Grässel, S. & Aszódi, A.) (Springer International Publishing, 2017).

23. Burstein, D., Bashir, A & Gray, M. L. MRI techniques in early stages of cartilage disease. *Investigative radiology* **35,** 622–38 (2000).

24. Kellgren, J. H. & Lawrence, J. S. Radiological assessment of osteo-arthrosis. *Annals of the rheumatic diseases* **16,** 494–502 (1957).

25. Petersson, I. F., Boegård, T, Saxne, T, Silman, A. J. & Svensson, B. Radiographic osteoarthritis of the knee classified by the Ahlbäck and Kellgren & Lawrence systems for the tibiofemoral joint in people aged 35-54 years with chronic knee pain. *Annals of the rheumatic diseases* **56,** 493–6 (1997).

26. Guermazi, A. *et al.* Compositional MRI techniques for evaluation of cartilage degeneration in osteoarthritis. *Osteoarthritis and Cartilage* **23,** 1639–1653 (2015).

27. Braun, H. J. & Gold, G. E. Advanced MRI of articular cartilage. *Imaging in Medicine* **3,** 541–555 (2011).

28. Hunter, D. J. *et al.* Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). eng. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* **19,** 990–1002 (2011).

29. Hunter, D. J. *et al.* The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston-Leeds Osteoarthritis Knee Score). *Annals of the Rheumatic Diseases* **67,** 206–211 (2008).

30. Peterfy, C. G. *et al.* Whole-organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis. *Osteoarthritis and Cartilage* **12,** 177–190 (2004).

31. Guermazi, A., Roemer, F. W., Haugen, I. K., Crema, M. D. & Hayashi, D. MRI-based semiquantitative scoring of joint pathology in osteoarthritis. *Nature Reviews Rheumatology* **9,** 236–251 (2013).

32. Stahl, R. *et al.* Osteoarthritis of the knee at 3.0 T: Comparison of a quantitative and a semi-quantitative score for the assessment of the extent of cartilage lesion and bone marrow edema pattern in a 24-month longitudinal study. *Skeletal Radiology* **40,** 1315–1327 (2011).

33. Deveza, L. *et al.* Trajectories of femorotibial cartilage thickness among persons with or at risk of knee osteoarthritis: development of a prediction model to identify progressors. *Osteoarthritis and Cartilage* **27,** 257–265 (2019).

34. Eckstein, F. *et al.* Non-invasive determination of cartilage thickness throughout joint surfaces using magnetic resonance imaging. *Journal of Biomechanics* **30,** 285–289 (1997).

35. Pelletier, J.-P. *et al.* Risk factors associated with the loss of cartilage volume on weight-bearing areas in knee osteoarthritis patients assessed by quantitative magnetic resonance imaging: a longitudinal study. *Arthritis Research & Therapy* **9,** 1–11 (2007).

36. Hayashi, D., Roemer, F. W. & Guermazi, A. Imaging of osteoarthritis - recent research developments and future perspective. *The British Journal of Radiology* **91,** 20170349 (2018).

37. Nieminen, M. T. *et al.* T 2 relaxation reveals spatial collagen architecture in articular cartilage: A comparative quantitative MRI and polarized light microscopic study. *Magnetic Resonance in Medicine* **46,** 487–493 (2001).

38. Zarins, Z. A. *et al.* Cartilage and meniscus assessment using T1rho and T2 measurements in healthy subjects and patients with osteoarthritis. *Osteoarthritis and Cartilage* **18,** 1408–1416 (2010).

39. Akella, S. V. S. *et al.* Proteoglycan-Induced Changes in T1ρ-Relaxation of Articular Cartilage at 4T. *Magnetic Resonance in Medicine* **46,** 419–423 (2001).

40. Duvvuri, U. *et al.* T1ρ-relaxation in articular cartilage: Effects of enzymatic degradation. *Magnetic Resonance in Medicine* **38,** 863–867 (1997).

41. MacKay, J. *et al.* Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis. *Osteoarthritis and Cartilage* **26,** 1140–1152 (2018).

42. Regatte, R. R., Akella, S. V., Lonner, J., Kneeland, J. & Reddy, R. T1ρ relaxation mapping in human osteoarthritis (OA) cartilage: Comparison of T1ρ with T2. *Journal of Magnetic Resonance Imaging* **23,** 547–553 (2006).

43. Hayashi, D., Roemer, F. W. & Guermazi, A. Recent advances in research imaging of osteoarthritis with focus on MRI, ultrasound and hybrid imaging. *Clinical and experimental rheumatology* **36 Suppl 1,** 43–52 (2018).

44. Bannuru, R. R. *et al.* OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. *Osteoarthritis and Cartilage* **27,** 1578–1589 (2019).

45. Fransen, M. *et al.* Exercise for osteoarthritis of the knee: A Cochrane systematic review. *British Journal of Sports Medicine* **49,** 1554–1557 (2015).

46. Coleman, J. L. *et al.* Diurnal variations in articular cartilage thickness and strain in the human knee. *Journal of Biomechanics* **46,** 541–547 (2013).

47. Waterton, J. C. *et al.* Diurnal variation in the femoral articular cartilage of the knee in young adult humans. *Magnetic Resonance in Medicine* **43,** 126–132 (2000).

48. Li, X. *et al.* Simultaneous Acquisition of T1ρ and T2 Quantification in Knee Cartilage : Repeatability and Diurnal Variation. *Journal of Magnetic Resonance Imaging* **39,** 1287–1293 (2014).

49. Sitoci, K. H., Hudelmaier, M. & Eckstein, F. Nocturnal changes in knee cartilage thickness in young healthy adults. *Cells Tissues Organs* **196,** 189–194 (2012).

50. Coleman, J. L. *et al.* Diurnal Variations in Articular Cartilage Thickness and Strain in the Human Knee. *J Biomech* **46,** 541–547 (2013).

51. Widmyer, M. R. *et al.* High body mass index is associated with increased diurnal strains in the articular cartilage of the knee. *Arthritis and Rheumatism* **65,** 2615–2622 (2013).

52. Taylor, K. A. *et al.* Activities of daily living influence tibial cartilage T1rho relaxation times. *Journal of Biomechanics* **82,** 228–233 (2019).

53. Eckstein, F *et al.* Effect of physical exercise on cartilage volume and thickness in vivo: MR imaging study. *Radiology* **207,** 243–8 (1998).

54. Eckstein, F., Tieschky, M., Faber, S., Englmeier, K. H. & Reiser, M. Functional analysis of articular cartilage deformation, recovery, and fluid flow following dynamic exercise in vivo. *Anatomy and Embryology* **200,** 419–424 (1999).

55. Eckstein, F., Lemberger, B., Stammberger, T., Englmeier, K. H. & Reiser, M. Patellar cartilage deformation in vivo after static versus dynamic loading. *Journal of Biomechanics* **33,** 819–825 (2000).

56. Hudelmaier, M *et al.* Age-Related Changes in the Morphology and Deformational Behavior of Knee Joint Cartilage. *Arthritis & Rheumatology* **44,** 2556–2561 (2001).

57. Liess, C., Lüsse, S., Karger, N., Heller, M. & Glüer, C. G. Detection of changes in cartilage water content using MRI T2-mapping in vivo. *Osteoarthritis and Cartilage* **10,** 907–913 (2002).

58. Eckstein, F. *et al.* In vivo cartilage deformation after different types of actvity and its dependence on physical training status. *Annals of the Rheumatic Diseases* **64,** 291–295 (2005).

59. Kessler, M. A., Glaser, C., Tittel, S., Reiser, M. & Imhoff, A. B. Volume changes in the menisci and articular cartilage of runners: An in vivo investigation based on 3-D magnetic resonance imaging. *American Journal of Sports Medicine* **34,** 832–836 (2006).

60. Kessler, M. A., Glaser, C., Tittel, S., Reiser, M. & Imhoff, A. B. Recovery of the menisci and articular cartilage of runners after cessation of exercise: Additional aspects of in vivo investigation based on 3-dimensional magnetic resonance imaging. *American Journal of Sports Medicine* **36,** 966–970 (2008).

61. Mosher, T. J., Liu, Y. & Torok, C. M. Functional cartilage MRI T2 mapping: evaluating the effect of age and training on knee cartilage response to running. *Osteoarthritis Cartilage* **18,** 358–364 (2010).

62. Farrokhi, S., Colletti, P. M. & Powers, C. M. Differences in patellar cartilage thickness, transverse relaxation time, and deformational behavior: A comparison of young women with and without patellofemoral pain. *American Journal of Sports Medicine* **39,** 384–391 (2011).

63. Niehoff, A. *et al.* Deformational behaviour of knee cartilage and changes in serum cartilage oligomeric matrix protein (COMP) after running and drop landing. *Osteoarthritis and Cartilage* **19,** 1003–1010 (2011).

64. Subburaj, K. *et al.* The acute effect of running on knee articular cartilage and meniscus magnetic resonance relaxation times in young healthy adults. *American Journal of Sports Medicine* **40,** 2134–2141 (2012).

65. Van Ginckel, A. & Witvrouw, E. Acute Cartilage Loading Responses After an In Vivo Squatting Exercise in People With Doubtful to Mild Knee Osteoarthritis: A Case-Control Study. *Physical Therapy* **93,** 1049–1060 (2013).

66. Sutter, E. G. *et al.* In Vivo Measurement of Localized Tibiofemoral Cartilage Strains in Response to Dynamic Activity. *The American Journal of Sports Medicine* **43,** 370–376 (2015).

67. Horng, A. *et al.* Topographic deformation patterns of knee cartilage after exercises with high knee flexion: an in vivo 3D MRI study using voxel-based analysis at 3T. *European Radiology* **25,** 1731–1741 (2015).

68. Lad, N. K. *et al.* Effect of normal gait on in vivo tibiofemoral cartilage strains. *Journal of Biomechanics* **49,** 2870–2876 (2016).

69. Gatti, A. A. *et al.* Acute changes in knee cartilage transverse relaxation time after running and bicycling. *Journal of Biomechanics* **53,** 171–177 (2017).

70. Liu, B. *et al.* In Vivo Tibial Cartilage Strains in Regions of Cartilage-to-Cartilage Contact and Cartilage-to-Meniscus Contact in Response to Walking. *American Journal of Sports Medicine* **45,** 2817–2823 (2017).

71. Chen, M. *et al.* The influences of walking, running and stair activity on knee articular cartilage: Quantitative MRI using T1 rho and T2 mapping. *PLoS ONE* **12,** 1–15 (2017).

72. Owusu-Akyaw, K. A. *et al.* A comparison of patellofemoral cartilage morphology and deformation in anterior cruciate ligament deficient versus uninjured knees. *Journal of Biomechanics* **67,** 78–83 (2018).

73. Ho, K.-Y., McClaren, J. & Sudweeks, S. Acute Effects of Walking on the Deformation of Femoral Articular Cartilage in Older Adults. *Journal of Geriatric Physical Therapy,* 1 (2018).

74. Collins, A. T. *et al.* Obesity alters the in vivo mechanical response and biochemical properties of cartilage as measured by MRI. *Arthritis Research & Therapy* **20,** 1–9 (2018).

75. Sutter, E. G. *et al.* Effects of Anterior Cruciate Ligament Deficiency on Tibiofemoral Cartilage Thickness and Strains in Response to Hopping. *American Journal of Sports Medicine* **47,** 96–103 (2019).

76. Nag, D., Liney, G. P., Gillespie, P. & Sherman, K. P. Quantification of T2 Relaxation Changes in Articular Cartilage With In Situ Mechanical Loading of the Knee. *Journal of Magnetic Resonance Imaging* **19,** 317–322 (2004).

77. Nishii, T., Kuroda, K., Matsuoka, Y., Sahara, T. & Yoshikawa, H. Change in Knee Cartilage T2 in Response to Mechanical Loading. *Journal of Magnetic Resonance Imaging* **28,** 175–180 (2008).

78. Souza, R. B. *et al.* The effects of acute loading on T1ρ and T2 relaxation times of tibiofemoral articular cartilage. *Osteoarthritis and Cartilage* **18,** 1557–1563 (2010).

79. Shin, C. S. *et al.* In vivo tibiofemoral cartilage-to-cartilage contact area of females with medial osteoarthritis under acute loading using MRI. *Journal of Magnetic Resonance Imaging* **34,** 1405–1413 (2011).

80. Cotofana, S. *et al.* In vivo measures of cartilage deformation: Patterns in healthy and osteoarthritic female knees using 3T MR imaging. *European Radiology* **21,** 1127–1135 (2011).

81. Subburaj, K. *et al.* Association of MR relaxation and cartilage deformation in knee osteoarthritis. *Journal of Orthopaedic Research* **30,** 919–926 (2012).

82. Souza, R. B. *et al.* Response of knee cartilage T1rho and T2 relaxation times to in vivo mechanical loading in individuals with and without knee osteoarthritis. *Osteoarthritis and Cartilage* **22,** 1367–1376 (2014).

83. Schoenbauer, E. *et al.* Cartilage evaluation with biochemical MR imaging using in vivo Knee compression at 3T-comparison of patients after cartilage repair with healthy volunteers. *Journal of Biomechanics* **48,** 3349–3355 (2015).

84. Lange, T., Knowles, B. R., Herbst, M., Izadpanah, K. & Zaitsev, M. Comparative T2 and T1ρ mapping of patellofemoral cartilage under in situ mechanical loading with prospective motion correction. *Journal of Magnetic Resonance Imaging* **46,** 452–460 (2017).

85. Lange, T. *et al.* Quantification of patellofemoral cartilage deformation and contact area changes in response to static loading via high-resolution MRI with prospective motion correction. *Journal of Magnetic Resonance Imaging,* 1–10 (2019).

86. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning (Adaptive Computation and Machine Learning Series)* 800 (MIT Press, 2017).

87. Hesamian, M. H., Jia, W., He, X. & Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging* **32,** 582–596 (2019).

88. Maas, A. L., Hannun, A. Y. & Ng, A. Y. *Rectifier nonlinearities improve neural network acoustic models* in *ICML Workshop on Deep Learning for Audio, Speech and Language Processing* (2013).

89. Dice, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **26,** 297–302 (1945).

90. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv preprint arXiv: 1502.03167v3* (2015).

91. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv: 1207.0580v1,* 1–18 (2012).

92. Jaccard, P. Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société vaudoise des sciences naturelles* **37,** 241–272 (1901).

93. Ronneberger, O., Fischer, P. & Brox, T. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* (eds Navab, N, Hornegger, J, Wells, W & Frangi, A) Lecture No, 234–241 (Springer, 2015).

94. Gaj, S., Yang, M., Nakamura, K. & Li, X. Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks. *Magnetic Resonance in Medicine,* 1–13 (2019).

95. Liu, F. SUSAN: segment unannotated image structure using adversarial network. *Magnetic Resonance in Medicine* **81,** 3330–3345 (2019).

96. Goodfellow, I. J. *et al.* Generative Adversarial Networks. *arXiv preprint arXiv:1406.2661v1,* 1–9 (2014).

97. Isola, P., Zhu, J. Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017,* 5967–5976 (2017).

98. Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R. & Venkatesan, R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design* Second Edi (Wiley Blackwell, 2014).

99. Chen, W. Errors in quantitative T1rho imaging and the correction methods. *Quantitative imaging in medicine and surgery* **5,** 583–91 (2015).

100. Li, Y., Zhao, F., Wang, Y. X., Ahuja, A. T. & Yuan, J. Study of magnetization evolution by using composite spin-lock pulses for T1ρ imaging. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS,* 408–411 (2012).

101. Wang, L. & Regatte, R. R. T1ρ MRI of human musculoskeletal system. *Journal of Magnetic Resonance Imaging* **41,** 586–600 (2015).

102. Charagundla, S. R., Borthakur, A., Leigh, J. S. & Reddy, R. Artifacts in T1ρ-weighted imaging: Correction with a self-compensating spin-locking pulse. *Journal of Magnetic Resonance* **162,** 113–121 (2003).

103. Witschey II, W. R. *et al.* Artifacts in T1ρ-Weighted Imaging: Compensation for B1 and B0 Field Imperfections. *Journal of Magnetic Resonance* **186,** 75–85 (2007).

104. Yuan, J. *et al.* Quantification of T1ρ relaxation by using rotary echo spin-lock pulses in the presence of B0 inhomogeneity. *Physics in Medicine and Biology* **57,** 5003–5016 (2012).

105. Becker, E. D., Feretti, J. A. & Farrar, T. C. Driven Equilibrium Fourier Transform Spectroscopy. A New Method for Nuclear Magnetic Resonance Signal Enhancement. *Journal of the American Chemical Society* **91,** 7784–7785 (1969).

106. Shoup, R. R., Becker, E. D. & Farrar, T. C. The Driven Equilibrium Fourier Transform NMR Technique: An Experimental Study. *Journal of Magnetic Resonance* **8,** 298–310 (1972).

107. Subburaj, K. *et al.* The Acute Effect of Running on Knee Articular Cartilage and Meniscus Magnetic Resonance Relaxation Times in Young Healthy Adults. *The American Journal of Sports Medicine* **40,** 2134–2141 (2012).

108. Nissi, M. J. *et al.* Proteoglycan and collagen sensitive MRI evaluation of normal and degenerated articular cartilage. *Journal of Orthopaedic Research* **22,** 557–564 (2004).

109. Cha, J. G. *et al.* Comparison of MRI T2 relaxation changes of knee articular cartilage before and after running between young and old amateur athletes. *Korean Journal of Radiology* **13,** 594–601 (2012).

110. Gatti, A. A. *et al.* Acute changes in knee cartilage transverse relaxation time after running and bicycling. *Journal of Biomechanics* **53,** 171–177 (2017).

111. Jordan, C. D. *et al.* Variability of CubeQuant T1rho, Quantitative DESS T2, and Cones Sodium MRI in Knee Cartilage. *Osteoarthritis Cartilage* **22,** 1559–1567 (2014).

112. Van Ginckel, A., Verdonk, P., Victor, J. & Witvrouw, E. Cartilage status in relation to return to sports after anterior cruciate ligament reconstruction. *American Journal of Sports Medicine* **41,** 550–559 (2013).

113. Oei, E. H. G., van Tiel, J., Robinson, W. H. & Gold, G. E. Quantitative Radiologic Imaging Techniques for Articular Cartilage Composition: Toward Early Diagnosis and Development of Disease-Modifying Therapeutics for Osteoarthritis. *Arthritis Care and Research* **66,** 1129–1141 (2014).

114. Li, X. *et al.* In vivo T1ρ and T2 mapping of articular cartilage in osteoarthritis of the knee using 3 T MRI. *Osteoarthritis and Cartilage* **15,** 789–797 (2007).

115. Jordan, C. D., Saranathan, M., Bangerter, N. K., Hargreaves, B. A. & Gold, G. E. Musculoskeletal MRI at 3.0 T and 7.0 T: A comparison of relaxation times and image contrast. *European Journal of Radiology* **82,** 734–739 (2013).

116. Chen, W., Takahashi, A. & Han, E. 3D Quantitative Imaging of T1rho and T2 (Abstract). *Proc Annual Meeting ISMRM* **19,** 231 (2011).

117. Van Rossom, S. *et al.* Knee cartilage thickness, T1ρ and T2 relaxation time are related to articular cartilage loading in healthy adults. *PLoS ONE* **12,** 1–16 (2017).

118. Levitt, M. H. & Freeman, R. Compensation for pulse imperfections in NMR spin-echo experiments. *Journal of Magnetic Resonance* **43,** 65–80 (1981).

119. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. W. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging* **29,** 196–205 (2010).

120. Treece, G. M., Prager, R. W. & Gee, A. H. Regularised marching tetrahedra: improved iso-surface extraction. *Comuters & Graphics* **23,** 583–598 (1999).

121. Treece, G., Gee, A., Mayhew, P. & Poole, K. High resolution cortical bone thickness measurement from clinical CT data. en. *Medical Image Analysis* **14,** 276–290 (2010).

122. Lassere, M. *et al.* Smallest detectable difference in radiological progression. *Journal of Rheumatology* **26,** 731–739 (1999).

123. Altman, D. & Bland, J. Measurement in Medicine : The Analysis of Method Comparison Studies. *The Statistician* **32,** 307–317 (1983).

124. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **1,** 307–310 (1986).

125. Wang, L. *et al.* T1rho MRI of menisci and cartilage in patients with osteoarthritis at 3T. *European Journal of Radiology* **81,** 2329–2336 (2012).

126. Hunter, D. J. & Eckstein, F. Exercise and osteoarthritis. *Journal of Anatomy* **214,** 197–207 (2009).

127. Blumenkrantz, G & Majumdar, S. Quantitative magnetic resonance imaging of articular. *European Cells and Materials* **13,** 76–86 (2016).

128. Benhamou, C. L. *et al.* Fractal analysis of radiographic trabecular bone texture and bone mineral density: Two complementary parameters related to osteoporotic fractures. *Journal of Bone and Mineral Research* **16,** 697–704 (2001).

129. Hunter, D. *et al.* Longitudinal validation of periarticular bone area and 3D shape as biomarkers for knee OA progression? Data from the FNIH OA Biomarkers Consortium. en. *Annals of the Rheumatic Diseases,* annrheumdis–2015–207602 (2015).

130. MacKay, J. W. *et al.* Association of subchondral bone texture on magnetic resonance imaging with radiographic knee osteoarthritis progression: data from the Osteoarthritis Initiative Bone Ancillary Study. *European Radiology* **28,** 4687–4695 (2018).

131. Neogi, T. *et al.* Magnetic resonance imaging-based three-dimensional bone shape of the knee predicts onset of knee osteoarthritis: Data from the osteoarthritis initiative. *Arthritis and Rheumatism* **65,** 2048–2058 (2013).

132. Wise, B. L. *et al.* Bone shape mediates the relationship between sex and incident knee osteoarthritis. *BMC Musculoskeletal Disorders* **19,** 1–9 (2018).

133. Bindernagel, M. *et al.* An articulated statistical shape model of the human knee. *Informatik aktuell,* 59–63 (2011).

134. Deniz, C. M. *et al.* Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks. *Scientific Reports* **8,** 1–14 (2018).

135. Lee, J. G., Gumus, S., Moon, C. H., Kwoh, C. K. & Bae, K. T. Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method. *Medical Physics* **41** (2014).

136. Ng, H., Ong, S., Foong, K., Goh, P. & Nowinski, W. Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm. *IEEE Southwest Symposium on Image Analysis and Interpretation,* 61–65 (2006).

137. Patel, F. K. & Singh, M. Segmentation of cartilage from knee MRI images using the watershed algorithm. *International Journal of Advance Research, Ideas and Innovations in Technology* **4,** 1727–1730 (2018).

138. Seim, H. *et al.* Model-based auto-segmentation of knee bones and cartilage in MRI data. *Medical Image Analysis for the Clinic: A Grand Challenge,* 215–223 (2010).

139. Shan, L., Zach, C., Charles, C. & Niethammer, M. Automatic atlas-based three-label cartilage segmentation from MR knee images. *Medical Image Analysis* **18,** 1233–1246 (2014).

140. Shrivastava, K., Gupta, N. & Sharma, N. Medical Image Segmentation using Modified K Means Clustering. *International Journal of Computer Applications* **103,** 12–16 (2014).

141. Swanson, M. *et al.* Semi-automated Segmentation to Assess the Lateral Meniscus in Normal and Osteoarthritic Knees. *Osteoarthritis Cartilage* **18,** 344–353 (2010).

142. Xia, Y. *et al.* Automated bone segmentation from large field of view 3D MR images of the hip joint. *Physics in Medicine and Biology* **58,** 7375–7390 (2013).

143. Zhou, L., Chav, R., Cresson, T., Chartrand, G. & De Guise, J. 3D knee segmentation based on three MRI sequences from different planes. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* **2016-Octob,** 1042–1045 (2016).

144. Shie, C. K., Chuang, C. H., Chou, C. N., Wu, M. H. & Chang, E. Y. Transfer representation learning for medical image analysis. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS,* 711–714 (2015).

145. Norman, B., Pedoia, V. & Majumdar, S. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. *Radiology* **288,** 177–185 (2018).

146. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. & Efros, A. A. Context Encoders: Feature Learning by Inpainting. *arXiv preprint arXiv:1604.07379v2,* 1–12 (2016).

147. Zhao, J. J., Mathieu, M. & LeCun, Y. Energy-based Generative Adversarial Network. *arXiv preprint arXiv: 1609.03126v4* (2017).

148. Rezaei, M. *et al.* A Conditional Adversarial Network for Semantic Segmentation of Brain Tumor. *BrainLes 2017: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer* **10670,** 241–252 (2017).

149. Dou, Q., Ouyang, C., Chen, C., Chen, H. & Heng, P. A. Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. *IJCAI International Joint Conference on Artificial Intelligence* **2018-July,** 691–697 (2018).

150. Huo, Y. *et al.* Adversarial synthesis learning enables segmentation without target modality ground truth. *Proceedings - International Symposium on Biomedical Imaging* **2018-April,** 1217–1220 (2018).

151. Chen, C., Dou, Q., Chen, H. & Heng, P. A. Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest X-ray segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11046 LNCS,** 143–151 (2018).

152. Heimann, T., Styner, M. & Warfield, S. K. *Segmentation of Knee Images : A Grand Challenge Segmentation of Knee Images :* 2010.

153. Ambellan, F., Tack, A., Ehlke, M. & Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. *Medical Image Analysis* **52,** 109–118 (2019).

154. *The Osteoarthritis Initiative, https://nda.nih.gov/oai/*

155. Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* **31,** 1116–1128 (2006).

156. Larobina, M. & Murino, L. Medical image file formats. *Journal of Digital Imaging* **27,** 200–206 (2014).

157. Regmi, K. & Borji, A. Cross-View Image Synthesis using Conditional GANs. *arXiv preprint arXiv:1803.03396v2* (2018).

158. Yang, D. *et al.* Automatic Liver Segmentation Using an Adversarial Image-to-Image Network. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017, Springer* **10435,** 507–515 (2017).

159. Li, C. & Wand, M. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. *arXiv preprint arXiv:1604.04382v1,* 1–17 (2016).

160. Long, J., Shelhamer, E. & Darrell, T. Fully Convolutional Adaptation Networks for Semantic Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 6810–6818 (2018).

161. Sørensen, T. J. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biologiske skrifter* **5,** 1–34 (1948).

162. Girshick, R. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision* **2015 Inter,** 1440–1448 (2015).

163. Felson, D. T. & Neogi, T. Osteoarthritis: Is It a Disease of Cartilage or of Bone? *Arthritis and Rheumatism* **50,** 341–344 (2004).

164. Zhou, Z., Zhao, G., Kijowski, R. & Liu, F. Deep convolutional neural network for segmentation of knee joint anatomy. *Magnetic Resonance in Medicine* **80,** 2759–2770 (2018).

165. Pedoia, V., Majumdar, S. & Link, T. M. Segmentation of joint and musculoskeletal tissue in the study of arthritis. *Magnetic Resonance Materials in Physics, Biology and Medicine* **29,** 207–221 (2016).

166. Fink, B. *et al.* Morphologic changes in the vastus medialis muscle in patients with osteoarthritis of the knee. *Arthritis and Rheumatism* **56,** 3626–3633 (2007).

167. Slemenda, C. *et al.* Quadriceps weakness and osteoarthritis of the knee. *Annals of Internal Medicine* **127,** 97–104 (1997).

168. Le Troter, A. *et al.* Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches. *Magnetic Resonance Materials in Physics, Biology and Medicine* **29,** 245–257 (2016).

169. Kemnitz, J. *et al.* Clinical evaluation of fully automated thigh muscle and adipose tissue segmentation using a U-Net deep learning architecture in context of osteoarthritic knee pain. *Magnetic Resonance Materials in Physics, Biology and Medicine* (2019).

170. Chaudhari, A. M. W., Briant, P. L., Bevill, S. L., Koo, S. & Andriacchi, T. P. Knee kinematics, cartilage morphology, and osteoarthritis after ACL injury. *Medicine and Science in Sports and Exercise* **40,** 215–222 (2008).

171. Messer, D. J., Shield, A. J., Williams, M. D., Timmins, R. G. & Bourne, M. N. Hamstring muscle activation and morphology are significantly altered 1–6 years after anterior cruciate ligament reconstruction with semitendinosus graft. *Knee Surgery, Sports Traumatology, Arthroscopy* **0,** 0 (2019).

172. Monu, U. *et al.* Cluster analysis of quantitative MRI T 2 and T 1ρ relaxation times of cartilage identifies differences between healthy and ACL-injured individuals at 3T. *Osteoarthritis and Cartilage* **25,** 513–520 (2017).

173. Paschos, N. K. Anterior cruciate ligament reconstruction and knee osteoarthritis. *World Journal of Orthopaedics* **8,** 212–217 (2017).

174. Lee, H., Hong, H. & Kim, J. *Anterior Cruciate Ligament Segmentation from Knee MR Images Using Graph Cuts with Geometric and Probabilistic Shape Constraints* in *Lee K.M., Matsushita Y., Rehg J.M., Hu Z. (eds) Computer Vision – ACCV 2012. Lecture Notes in Computer Science, vol 7725. Springer, Berlin, Heidelberg* (2013), 305 –315.

175. Paproki, A. *et al.* Automated segmentation and T2-mapping of the posterior cruciate ligament from MRI of the knee: Data from the osteoarthritis initiative. *Proceedings - 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI),* 424–427 (2016).

176. Mallya, Y., J., V., M. S., V., Venugopal, V. K. & Mahajan, V. Automatic delineation of anterior and posterior cruciate ligaments by combining deep learning and deformable atlas based segmentation. *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging.* **10953** (2019).

177. Hunter, D. J. & Bierma-Zeinstra, S. Osteoarthritis. *The Lancet* **393,** 1745–1759 (2019).

178. Hunter, D. J., Schofield, D. & Callander, E. The individual and socioeconomic impact of osteoarthritis. *Nature Reviews Rheumatology* **10,** 437–441 (2014).

179. Williams, T. G. *et al.* Anatomically corresponded regional analysis of cartilage in asymptomatic and osteoarthritic knees by statistical shape modelling of the bone. *IEEE Transactions on Medical Imaging* **29,** 1541–1559 (2010).

180. Norouzi, A. *et al.* Medical Image Segmentation Methods, Algorithms, and Applications. *IETE Technical Review* **31,** 199–213 (2014).

181. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation* in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (eds Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G. & Wells, W.) **9901 LNCS** (Springer International Publishing, Cham, 2016), 424–432.

182. Milletari, F., Navab, N. & Ahmadi, S. A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings - 4th International Conference on 3D Vision, 3DV 2016,* 565–571 (2016).

183. Kessler, D. A. *et al.* The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs. *Computerized Medical Imaging and Graphics* **86** (2020).

184. Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39,** 2481–2495 (2017).

185. Wirth, W. *et al.* Accuracy and longitudinal reproducibility of quantitative femorotibial cartilage measures derived from automated U-Net-based segmentation of two different MRI contrasts: data from the osteoarthritis initiative healthy reference cohort. *Magnetic Resonance Materials in Physics, Biology and Medicine* (2020).

186. Kessler, D. A. *et al.* Effectively Measuring Exercise-Related Variations in T1ρ and T2 Relaxation Times of Healthy Articular Cartilage. *Journal of Magnetic Resonance Imaging* **52,** 1753–1764 (2020).

187. McCormick, M., Liu, X., Jomier, J., Marion, C. & Ibanez, L. ITK: enabling reproducible research and open science. *Frontiers in Neuroinformatics* **8,** 1–11 (2014).

188. Kingma, D. P. & Ba, J. L. Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations (ICLR), San Diego, USA,* 1–15 (2015).

189. Taghanaki, S. A. *et al.* Combo loss: Handling input and output imbalance in multi-organ segmentation. *Computerized Medical Imaging and Graphics* **75,** 24–33 (2019).

190. Hu, Y., Chen, X., Wang, S., Jing, Y. & Su, J. Subchondral bone microenvironment in osteoarthritis and pain. *Bone Research* **9,** 20 (2021).

191. Findlay, D. M. & Kuliwaba, J. S. Bone–cartilage crosstalk: a conversation for understanding osteoarthritis. *Bone Research* **4,** 16028 (2016).

192. Goldring, S. R. Alterations in periarticular bone and cross talk between subchondral bone and articular cartilage in osteoarthritis. *Therapeutic Advances in Musculoskeletal Disease* **4,** 249–258 (2012).

193. Van Helvoort, E. M. *et al.* Cohort profile: The Applied Public-Private Research enabling OsteoArthritis Clinical Headway (IMI-APPROACH) study: a 2-year, European, cohort study to describe, validate and predict phenotypes of osteoarthritis using clinical, imaging and biochemical mark. *BMJ Open* **10,** e035101 (2020).

194. Desai, A. D. *et al.* The International Workshop on Osteoarthritis Imaging Knee MRI Segmentation Challenge: A Multi-Institute Evaluation and Analysis Framework on a Standardized Dataset. *Radiology: Artificial Intelligence* **3,** e200078 (2021).

195. Karimi, D., Warfield, S. K., Gholipour, A. & Member, S. Transfer Learning in Medical Image Segmentation: New Insights from Analysis of the Dynamics of Model Parameters and Learned Representations. *Artificial Intelligence In Medicine* **116,** 102078 (2020).

196. Turmezei, T. *et al.* Test-restest repeatability and reproducibility of 3D joint space mapping at the ankle using weight bearing cone beam computed tomography. *Osteoarthritis and Cartilage* **29,** S317–S318 (2021).

197. Turmezei, T. D. *et al.* Quantitative Three-dimensional Assessment of Knee Joint Space Width from Weight-bearing CT. *Radiology* **299,** 649–659 (2021).

198. Treece, G. & Gee, A. Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. *Medical Image Analysis* **20,** 249–264 (2015).

199. Turmezei, T. D., Treece, G. M., Gee, A. H., Fotiadou, A. F. & Poole, K. E. S. Quantitative 3D analysis of bone in hip osteoarthritis using clinical computed tomography. *European Radiology* **26,** 2047–2054 (2016).

200. Kazhdan, M. & Hoppe, H. Screened poisson surface reconstruction. *ACM Transactions on Graphics* **32,** 1–13 (2013).

201. Zhou, Q.-Y., Park, J. & Koltun, V. Open3D: A Modern Library for 3D Data Processing (2018).

202. Moré, J. J. in *Numerical Analysis. Lecture Notes in Mathematics* (ed Watson, G.) 105–116 (Springer, Berlin, Heidelberg, 1978).

203. Gunz, P., Mitteroecker, P. & Bookstein, F. L. in *Modern Morphometrics in Physical Anthropology* 73–98 (Kluwer Academic Publishers-Plenum Publishers, New York, 2004).

204. Turmezei, T. D. *et al.* Quantitative 3D imaging parameters improve prediction of hip osteoarthritis outcome. *Scientific Reports* **10,** 4127 (2020).

205. Van Ginckel, A. *et al.* Functional adaptation of knee cartilage in asymptomatic female novice runners compared to sedentary controls. A longitudinal analysis using delayed Gadolinium Enhanced Magnetic Resonance Imaging of Cartilage (dGEMRIC). *Osteoarthritis and Cartilage* **18,** 1564–1569 (2010).

206. Hochberg, M. C. *et al.* Effect of Intra-Articular Sprifermin vs Placebo on Femorotibial Joint Cartilage Thickness in Patients With Osteoarthritis. *JAMA* **322,** 1360 (2019).

207. Jansen, M. P., Mastbergen, S. C., MacKay, J. W., Turmezei, T. D. & Lafeber, F. Knee joint distraction results in MRI cartilage thickness increase up to ten years after treatment. *Rheumatology* (2021).

208. Ma, D. *et al.* Magnetic Resonance Fingerprinting. *Nature* **495,** 187–192 (2013).