Title
Computer Vision for Morphological Evaluation of Musculoskeletal Disorders in Magnetic Resonance Imaging

Author
Gao, Kenneth

Publication Date
2023

Peer reviewed|Thesis/dissertation

Computer Vision for Morphological Evaluation of Musculoskeletal Disorders in Magnetic Resonance Imaging

by

Kenneth Gao

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:
Sharmila Majumdar
ED3F278527E84B0...

Sharmila Majumdar

Chair

DocuSigned by:
Thomas M. Link
D07C...

Thomas M. Link

DocuSigned by:
Valentina Pedoia
D8C...

Valentina Pedoia

DocuSigned by:
Grace O'Connell
B03D...

Grace O'Connell

DocuSigned by:
Yang Yang
731E1D96DE5B471...

Yang Yang

Committee Members

Acknowledgements

To my adviser, Sharmila Majumdar, you taught me to aim high and challenge the status quo. You gave me confidence when I needed it and you reined me in when my imagination wandered. Each of our conversations led to action, and without your contagious ambition, I would not be where I am today.

To Valentina Pedoia, your enthusiasm shines wherever you go. You are unafraid to dive headfirst into new, unfamiliar topics. Among all that you taught me, what I value most is your ability to stimulate excitement in others. Thank you for continuing to believe in me.

To my labmates, past and present—Erin Argentieri, Upasana Bharadwaj, Rupsa Bhattacharjee, Francesco Calivá, Felix Gassert, Misung Han, Madeline Hess, Gabrielle Hoyer, Claudia Iriondo, Jenny Lee, Felix Liu, Johanna Luitjens, Alejandro Morales Martinez, Susanna Morin, Cameron Nosrat, Eugene Ozhinsky, Koren Roach, Rafeek Thahakoya, Radhika Tibrewala, Aniket Tolpadi, and Michelle Tong—thank you for the scientific discussions, casual chats, and spontaneous adventures. Your work is inspiring, and I look forward to following your achievements.

To Cynthia Chin, Thomas Link, and Richard Souza, you guided the clinical relevance of my work and invigorated my love for health and medicine.

To my qualifying exam committee members—Sri Nagarajan, Cynthia Chin, Janine Lupo, and Grace O'Connell—thank you for challenging me to think beyond the scope of my proposal. You guided me to think carefully of theory, clinical impact, and pragmatism.

To my other mentors and colleagues in the UCSF Department of Radiology and Biomedical Imaging—Emma Bahroos, Michael Evans, Peder Larson, Youngho Seo, Daniel Vigneron, Henry VanBrocklin, Duan Xu, Yang Yang, and many others—though our work never intersected, you taught me the principles of being a scientist.

To my peers and the UC Berkeley-UCSF Joint Graduate Program in Bioengineering, I am consistently amazed by your rigor, in all areas of science, mentorship, and inclusivity. Thank you for being a welcoming and progressive community.

To my parents, through my childhood you guided me to be the best person I can be. Your love, support, and sacrifices have shaped me into the person I am today. This dissertation is dedicated to you.

Finally, to my wife, Stephanie, you make me a complete person. Through an arduous 5-year span of this PhD, we were married, survived a pandemic, and raised a spoiled puppy. I know that we will meet every challenge successfully, together. Thank you for being my foundation.

Kenneth Gao

May 2023

Abstract

Computer Vision for Morphological Evaluation of

Musculoskeletal Disorders in Magnetic Resonance Imaging

by Kenneth T. Gao

With the aging of the general population, musculoskeletal (MSK) diseases have moved to the forefront of healthcare concerns and are the leading causes of disability globally. Noninvasive imaging is routinely utilized in the clinic to diagnose and monitor onset and progression of MSK conditions. However, due to the qualitative nature of imaging assessments and increasing labor costs of evaluating advanced imaging modalities, there is a crucial need for automatic quantitative approaches. In this dissertation, we explore the development of computer vision techniques for extracting morphological features associated with low back pain and knee osteoarthritis, two of the most prevalent and debilitating MSK conditions.

We begin by addressing the costs of image annotation via automation with deep learning. More specifically, we developed convolutional neural networks for two purposes: (1) to semantically segment various tissues, allowing for geometric tissue characterization, and (2) to detect and localize lesions and abnormalities. Then, leveraging these models for feature extraction, we harmonized tissue geometries in 3D Euclidean space using atlas-based registration to identify tissue shapes predisposed to disease onset. These techniques were applied to both large-scale and small, limited datasets, demonstrating the utility of computer vision techniques for morphological evaluation in a data-driven, exploratory manner.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACL	anterior cruciate ligament
AI	artificial intelligence
ANCOVA	analysis of covariance
ANN	artificial neural network
BMI	body mass index
CI	confidence interval
cLBP	chronic low back pain
CSA	cross-sectional area
DESS	dual-echo steady-state
DESS-we	dual-echo steady-state with selective water excitation
DL	deep learning
ECM	extracellular matrix
ES	erector spinae
ETL	echo train length
FSE	fast spin-echo
GAG	glycosaminoglycan
GE	General Electric
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IVD	intervertebral disc
IVDH	intervertebral disc height
KL	Kellgren-Lawrence

LBP	low back pain
LFC	lateral femoral condyle
LT	lateral tibia
MAPSS	magnetization-prepared angle-modulated partitioned *k*-space spoiled gradient echo snapshots
MC	Modic change
MFC	medial femoral condyle
ML	machine learning
MOAKS	MRI Osteoarthritis Knee Score
MR	magnetic resonance
MRI	magnetic resonance imaging
MT	medial tibia
NBA	National Basketball Association
NCAA	National Collegiate Athletic Association
OA	osteoarthritis
OAI	Osteoarthritis Initiative
PAT	patella
PC	principal component
PCA	principal component analysis
PCL	posterior cruciate ligament
PD	proton density
PG	proteoglycan
RF	radiofrequency

ROI	region-of-interest
QL	quadratus lumborum
SD	standard deviation
SNR	signal-to-noise ratio
SSM	statistical shape modeling
TE	echo time
TR	repetition time
TRO	trochlea
TSL	time of spin lock
UCSF	University of California, San Francisco
VB	vertebral body
VBR	voxel-based relaxometry

Chapter 1

Introduction

In all aspects of biology, there is a deep connection between form (“morphology”) and function. These relationships span across every level of human biology. To paint an example, we will look towards knee osteoarthritis (OA), a painful degenerative disease characterized by the degradation of the hyaline cartilage tissue that protects and allows for articulation of the joint. While there are many proposed mechanisms for the development of OA, morphology has been established as a strong correlate. On the tissue-level, the two hallmark imaging features of the disease are morphological: (1) cartilage loss, and (2) growth of bony osteophytes. Within the cartilage, the structure of the dense extracellular matrix informs our understanding of how force is transmitted across the cartilage layers during knee flexion and extension. And the organization of collagen fibers, chondrocytes, and proteoglycan complexes that form the cartilage are critical to the tissue’s mechanical properties. In this manuscript, we will broach the association of these various levels of morphometry to disease, with a particular focus on tissue geometry.

Noninvasive imaging has been a mainstay in medical diagnostics and the assessment of MSK diseases. Radiographs (x-rays) have been the defining method of diagnosis of knee OA; however, radiography is not able to depict the articular cartilage and is limited as a 2D projection modality. In contrast, magnetic resonance imaging (MRI) has proven capable of high resolution, volumetric delineation of various soft tissues and is now widely used in research and clinical trial settings. However, the time and labor involved in evaluating advanced magnetic resonance (MR) sequences are dramatically higher than that of plain film radiographs, particularly when accounting for the complexity and flexibility of MRI technology that has introduced high variability in image contrast.

We address this challenge with deep learning (DL), a recently coined term for a family of machine learning (ML) techniques that has seen dramatic leaps in outcomes, particularly in natural imaging (*e.g.,* photos taken with a digital camera). The transfer of these technologies to the medical imaging field has been studied with necessary rigor and with careful regard to its application. In the efforts discussed here, we investigate the development of DL-based segmentation to extract tissue morphometrics and its application in experimental studies to quantify group-level characteristics of populations at risk for low back pain (LBP) and knee OA.

1.1 Outline

In Chapter 2, we summarize some background information in areas of disease epidemiology and mechanisms, imaging physics, and DL-based representation learning. Chapter 3 describes the implementation of DL models for tissue segmentation in the lumbar spine. We expand upon these DL models in Chapter 4 to demonstrate how segmentation models can be utilized for detecting abnormal spinal pathologies.

Subsequent chapters shift from LBP to knee OA, in which we spatially align DL-derived tissue segmentations between subjects to understand variations in tissue geometry and composition. Chapter 5 specifically investigates early changes of cartilage composition and bone shape in a high-risk population—collegiate basketball players. Meanwhile, Chapter 6 demonstrates a “big data” approach to understanding the role of shape of the meniscus tissue in OA incidence, using thousands of MRIs to fit a statistical shape model. Finally, in Chapter 7 we summarize the contributions of this dissertation and several future paths of inquiry based upon these works.

Chapter 2

Background

2.1 Landscape of MSK Conditions

The MSK system encompasses the tissues involved in locomotion. Most commonly, we first think of the bones that give the body structure and the muscles that move them. Then, we consider secondary structures, such as tendons, ligaments, and cartilage, which play essential roles in facilitating forces and stability throughout the system. During the span of a lifetime, these components of the MSK system are vulnerable to injury and degeneration. Some tissues are adept at repair. However, it is often the secondary connective tissues in which healing is slow or nonexistent due to their hypocellular and hypovascular nature. Chronic MSK conditions account for the highest forms of disability and the most frequent visits to physicians' offices [1] and, as humans continue to live longer, these diseases will become an increasingly larger socioeconomical burden. In the following sections, we summarize the etiology, epidemiology, and diagnostic and management strategies for LBP and knee OA, two of the most debilitating forms of MSK disease.

2.1.1 Low Back Pain

2.1.1.1 Epidemiology and Impact

LBP is the highest cause of disability globally [2], affecting 540 million people in 2015. In the United States, nearly 29% of the population age 18 years and older reported suffering from LBP [1]. The highest rates were found in those between 45-64 years of age, though incidence is extremely common in people of all ages. Women and non-Hispanic white ethnic group were subpopulations with relatively higher prevalence than their peers.

In 2013, 1 in 4 persons in the United States visited a healthcare professional for back pain, accounting for 6.4% of all healthcare visits [1]. Since 1998, the rate of physician visits for back pain has been rapidly increasing, with the large majority of the increase attributed to LBP (**Figure 2.1**). Estimates for direct healthcare costs in the United States vary widely; most reports range between $50-100 billion in annual costs [3].

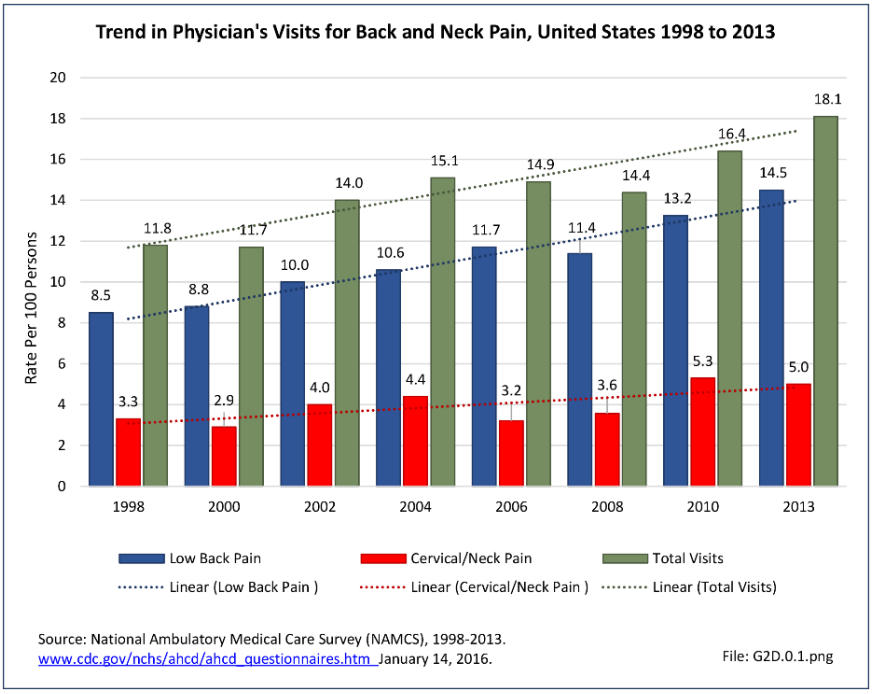

Figure 2.1 Rate of physician visits for back pain in the United States between 1998-2013. The rapid and steady increase is driven predominantly by LBP. Image source: [1].

Typically, indirect costs of this condition are higher than direct costs [4]. Many people with LBP do not seek care [5], particularly those with minor forms of disease. Severe cases account for 77% of all LBP-related disability, amounting up to 46.5 million years lived with disability worldwide [6]. This can lead to profound economic effects. Beyond the repercussions of lost workdays, people with the disorder accumulate substantially less wealth than those without. LBP is the leading reason for workers to retire prematurely [7]. For the individual, living with LBP leads to concerns for seeking diagnosis and treatment, regaining previous levels of health and function, or otherwise coping with disease [8].

2.1.1.2 Clinical Presentation, Diagnosis, and Pathophysiology

The presentation of LBP is often accompanied by concurrent pain in other locations and comorbidities. Psychological, social, and biophysical factors can all contribute to pain mechanisms. Smoking [9], obesity [10], and low levels of physical activity [11] have been associated with higher incidence. Past pain episodes and living with other chronic conditions, such as asthma, headache, and diabetes, are also associated with increased odds for future back pain [12]. Considering its multifactorial nature, LBP is challenging to evaluate in isolation.

Acute cases can be caused by physical or psychosocial factors [13]. One-third of acute episodes cannot be traced to a specific trigger by the patients [14] and the large majority of patients that can discern the cause attribute pain to the act of lifting [15].

It is important to note that LBP is a symptom that can result from various abnormalities and diseases [16]. In the case of chronic LBP, for the vast majority of people, the nociceptive source cannot be identified [17]. Diagnostic triage generally consists of a

history and physical examination to rule out possibilities of red flags. These include vertebral fractures, inflammatory arthritis, cauda equina, malignancy, and infection [16], [17]. Further diagnostic work-up may be warranted if a specific disease is suspected. Imaging, for example, is urgently recommended for possible cauda equina, infection, or strong suspicion of cancer. Blood tests are suggested for the latter two cases. Neurological cases, such as radicular pain, radiculopathy, or spinal canal stenosis, rely on clinical findings of leg pain or pain during lumbar flexion. If the above disorders are ruled out, then the case is designated as non-specific.

Several lumbar features have been identified as contributors to pain. Degenerative changes to the intervertebral discs (IVD) is one such morphology associated with LBP [18]. Providing structural support and allowing for locomotion of the spinal column, IVDs are cushion-shaped tissues that sit between each vertebra. They consist of three major components: (1) the nucleus pulposus, a gel-like structure that consists mostly of water, type II collagen, and proteoglycans, (2) the annulus fibrosus, which is composed of highly organized concentric rings of collagen, elastic fibers, glycoproteins, and proteoglycans, and (3) the endplates that anchor the disc to the neighboring vertebrae. The nucleus pulposus sits in the center of each disc and distributes hydraulic pressure via its high water content, while the annulus fibrosus surrounds the nucleus pulposus to provide the necessary structural foundation. With age, disc desiccation is common due to imbalances in anabolic synthesis and catabolic breakdown of ECM [19]. This results in a progressive replacement of the nucleus pulposus with fibrotic tissue and a loss of compressive resistance. To compensate, the annulus fibrosus endures additional force and undergoes remodeling and morphological changes [20]. These may include prolapse,

herniation, or loss in disc height, which may accompany changes in the vertebrae, such as spondylolysis, spondylolisthesis, or degenerative endplate (Modic) changes. The interplay between IVD degeneration, abnormal morphologies, and biomechanical instability is a strong culprit for pain and has been studied heavily [18]–[22].

Abnormalities of the vertebral endplates, the interface between IVDs and vertebral bodies, have also been linked to pain. The endplates play an essential role in nutrient transport between vertebral capillaries and the disc, providing glucose and oxygen to the cells in the disc and allowing for waste to be removed [22]. Upon skeletal maturity, endplates undergo remodeling and mineralization, thus becoming limited in its function of nutrient exchange. Modic changes (MC) are one of the initial systems used to describe common abnormalities of the endplate and vertebral body [23]. Three types of lesions are described, indicating fibrovascular replacement, fatty marrow replacement, and sclerosis, and are visualized using MRI. MCs are described in further detail as the primary focus of Chapter 4.

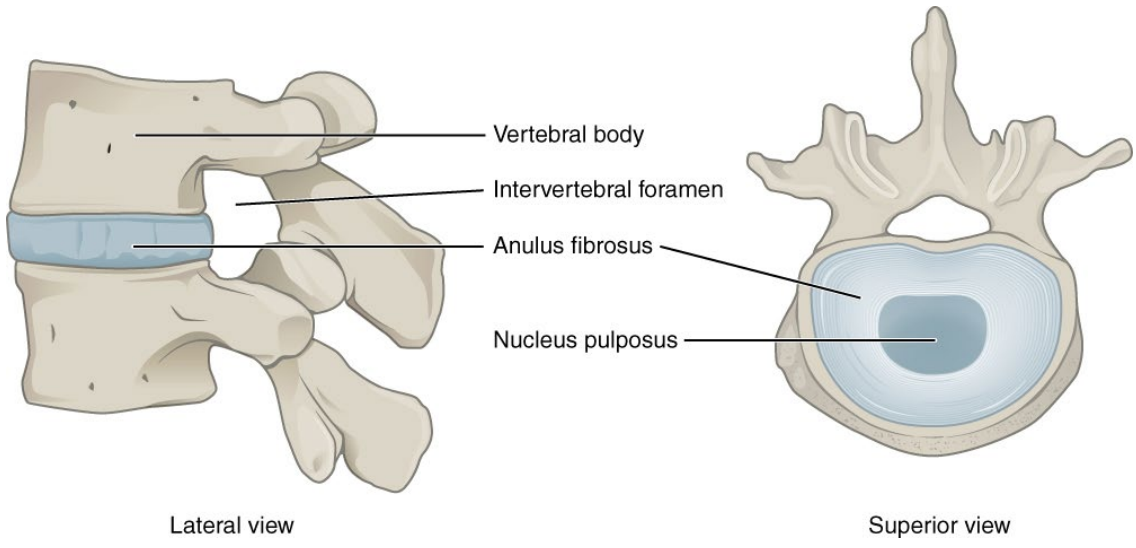

Figure 2.2 Schematic of the vertebrae and IVD. The IVD sits between the vertebral bodies, allowing for flexibility while providing structure. Image source: Jmarchn, Wikipedia.

While the IVD and vertebral body interface is the primary connection between each vertebra, the facet joints in the posterior aspect of the vertebral column are additional regions of articulation and are likewise vulnerable to degeneration. Characterized by loss of cartilage which facilitates the articulation of the superior and inferior spinal processes, facet joint osteoarthritis is strongly associated with disc degeneration due to the interdependency between the facets and disc in the transmission of force [24], [25].

The function of the vertebral column relies on the coexistence of the abovementioned structures. It is evident that deterioration of one tissue is closely linked to degeneration of its neighbors [21], [26], [27]. However, the association of LBP with these findings is inconsistent, both in terms of sensitivity (*i.e.,* presence of imaging abnormalities in patients with pain) and specificity (*i.e.,* presence of abnormalities in asymptomatic patients) [28], [29]. It is crucial to move towards a holistic approach for evaluating LBP to improve outcomes [30], [31].

2.1.1.4 Prevention and Management

Prevention strategies revolve around minimizing risk factors. Exercise has proven effective for improving general health and strengthening spinal stability [32]. Education alone did not improve outcomes; however, in conjunction with exercise, proved to be more effective than exercise in isolation [33]. Other forms of prevention, such as lifting devices, braces, ergonomic furniture, or shoe insoles, have yet to be conclusively investigated or show minimal effect to prevent LBP.

Due to unknown pathoanatomical sources of pain in nonspecific LBP, management focuses on reducing pain while minimizing further risk to mobility. Education, analgesics, and moderate activity are recommended when possible [17].

Pharmacological approaches (*e.g.,* NSAIDs, paracetamol, opioids) help ease pain and triggers. Non-pharmacological options are available as well, including physical therapy, massage, and acupuncture. In cases where the pain source is identifiable, targeted interventions can be undergone [34]. Targeted corticosteroid injections have demonstrated modest, short-lived efficacy for radicular pain and antidepressants and anticonvulsants aid in treating neuropathic pain. Surgery can be an option to address morphological defects such as degenerated discs, fracture, or stenosis. Ultimately, there is strong consensus that multimodal treatment is most effective, leveraging education, pharmacology, physical therapy, and possible surgical intervention.

2.1.2 Osteoarthritis

OA has traditionally been considered a wear-and-tear disease, characterized by mechanical degradation of cartilage and narrowing of the joint space. However, joints are comprised of many different tissues and experience varying forms of degenerative and pathological changes. As we understand more about OA, we find that phenotyping this complex disease is challenging and multifaceted [35]. Currently, there is no single endpoint for clinical trials targeting OA due to its heterogeneity, making it a critical public health concern.

2.1.2.1 Epidemiology and Impact

The global prevalence of hip and knee OA was estimated to be 3.7%, affecting 303.1 million people [36]. Incidence of knee OA increases with age due to accumulated exposure to risk factors and age-related biochemical changes [37], [38]. Beyond age, women, African Americans, and individuals with lower education status were three demographic groups in the United States found to be at higher risk. Other risk factors for

OA are closely tied to lifestyle, including body mass index (BMI), physical activity, and past injury. Expansion of the general population, increased lifespans, and higher rates of obesity have rapidly increased the prevalence of OA. According to figures from the Global Burden of Disease Study 2019, prevalence increased by 113.25% from 1990 to 2019 [39]. The knee is the most common site of OA, accounting for 60.6% of all cases globally, and will be the focus of the remainder of this section.

2.1.2.3 Clinical Presentation, Diagnosis, and Pathophysiology

Diagnosis of knee OA can be categorized into two forms: (1) symptomatic OA, and (2) radiographic OA. Symptomatic OA indicates that individuals experience pain, aching, and stiffness and is evaluated through a history and physical examination with load bearing and joint motion [40]. Symptoms are generally worsened with movement or after long periods of rest. Radiographic OA, diagnosed through imaging, has long been the reference standard. The most common definition is the Kellgren-Lawrence (KL) grading scheme [41]. Using posteroanterior radiographs in weightbearing position, the knee joints are assessed for osteophyte growth, sclerosis, and joint space narrowing indicative of cartilage loss and meniscal degeneration (**Figure 2.3**). There is inconsistent evidence of the correlation between degree of radiographic OA and pain [42]–[44] and evaluation of OA should not solely rely on radiography.

A major limitation of radiography is its inability to delineate cartilage. Providing resistance to external forces and allowing for articulation of the bones, cartilage is composed of mostly type II collagen, which provides a structural organization to the tissue [45] (**Figure 2.4**). Proteoglycans and glycosaminoglycans (GAGs) embedded in

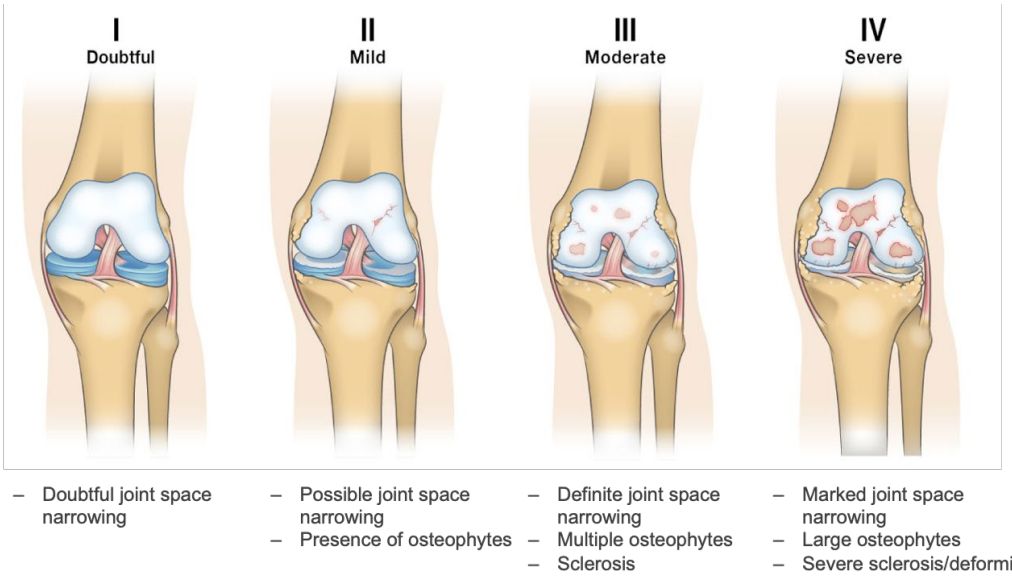

Figure 2.3 Diagram of KL grading for radiographic OA. Image adapted from *Spring Loaded Technologies*.

the collagen matrix attract water into the cartilage through its net negative charge, providing compressive force resistance. The composition of the tissue is regulated solely by chondrocytes, which respond to a variety of factors including mechanical, biochemical, and immune factors. In the degenerative process, compositional homeostasis is disrupted, leading to loss of structural integrity. Inflammatory cytokines and proteases are the main contributors to catabolic activities that degrade the cartilaginous ECM and predispose the tissue for mechanical failure [46]. Consequently, the altered ECM induce biomechanical repercussions. In early phases of OA, increased chondrocyte activity results in thicker cartilage thickness, softer ECM, and lower shear modulus. Later stages of disease are characterized by stiffened and thinning cartilage, due to higher prevalence of collage crosslinks and decreased proteoglycan content, causing greater load to shift to the underlying subchondral bone. This feedback loop of biochemical and mechanical changes is core to the destructive OA process.

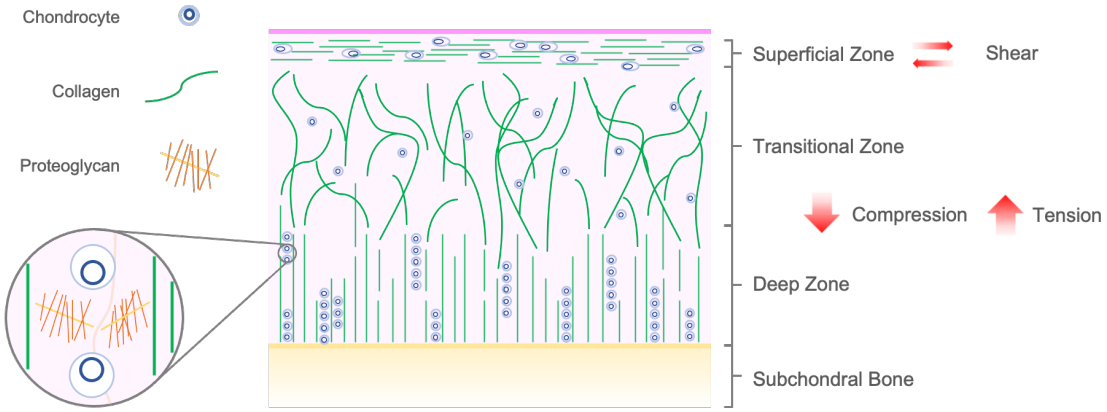

Figure 2.4 Representation of cartilage composition. Chondrocytes, the cellular component, manage the ECM, which primarily comprises of type II collagen and proteoglycan to provide structure and force resistance.

2.1.2.4 Prevention and Management

The two controllable risk factors for knee OA most often addressed in preventative strategies are obesity and knee injury [38]. Thus far, only one randomized control trial of knee OA prevention exists, which found that a 5 *kg* or 5% weight loss in the first year led to a threefold reduction in incidence of symptomatic knee OA and 2.5 times reduction in radiographic OA incidence [47]. Additional investigations of prevention techniques are warranted.

In individuals with symptomatic OA, pain management is similar to that for LBP. Exercise [48], physical therapy, and weight loss [49] aim to reduce force exerted on the knees and improve joint motion. Analgesics like paracetamol and NSAIDs are commonly recommended, and for non-responders, intra-articular corticosteroids can be explored for short-term relief [50], though effectiveness is still unclear [51].

Direct interventions for knee OA attempt to reduce loading of affected areas. Knee osteotomy is a surgical technique that improves alignment and shifts weight to healthier tissue by reshaping the tibial bone. Arthroscopic surgery can be considered to repair and

recontour damaged tissue such as the meniscus. However, rigorous evidence regarding its efficacy is lacking and older patients are typically not good candidates [52], [53]. Knee arthroplasty (joint replacement) has shown to be effective in late stage disease, when all other conservative interventions have failed [54]. However, there are many disadvantages: (1) the effective lifetime of prostheses are limited, (2) up to 25% of patients who undergo surgery continue to endure pain and disability one-year later [55], and (3) total knee replacements are associated with higher rates of serious adverse events than non-surgical options [56]. Advanced stages of disease are limited in interventional options. Thus, researchers are shifting focus to prevention and detection of early-stage disease.

2.2 MRI of Low Back Pain and Osteoarthritis

While imaging may be a weak predictor for pain, it serves as a vital tool for identifying abnormal morphology associated with disease. Imaging techniques continue to advance, pushing towards quantitative and more precise depiction of the underlying tissue for better understanding of pain mechanisms and disease outcomes. MRI, in particular, has become increasingly utilized for its powerful ability to delineate soft tissue in high resolution without exposure to ionizing radiation. In this section, we will introduce fundamentals in MRI signals. Then, we will compare the utility of conventional MRI used in the clinic and the advantages and disadvantages of advanced MRI techniques.

2.2.2 MRI Physics

Magnetic resonance was first known as nuclear magnetic resonance. By breaking down the individual components of its alias, we can more easily understand the physics behind this imaging system.

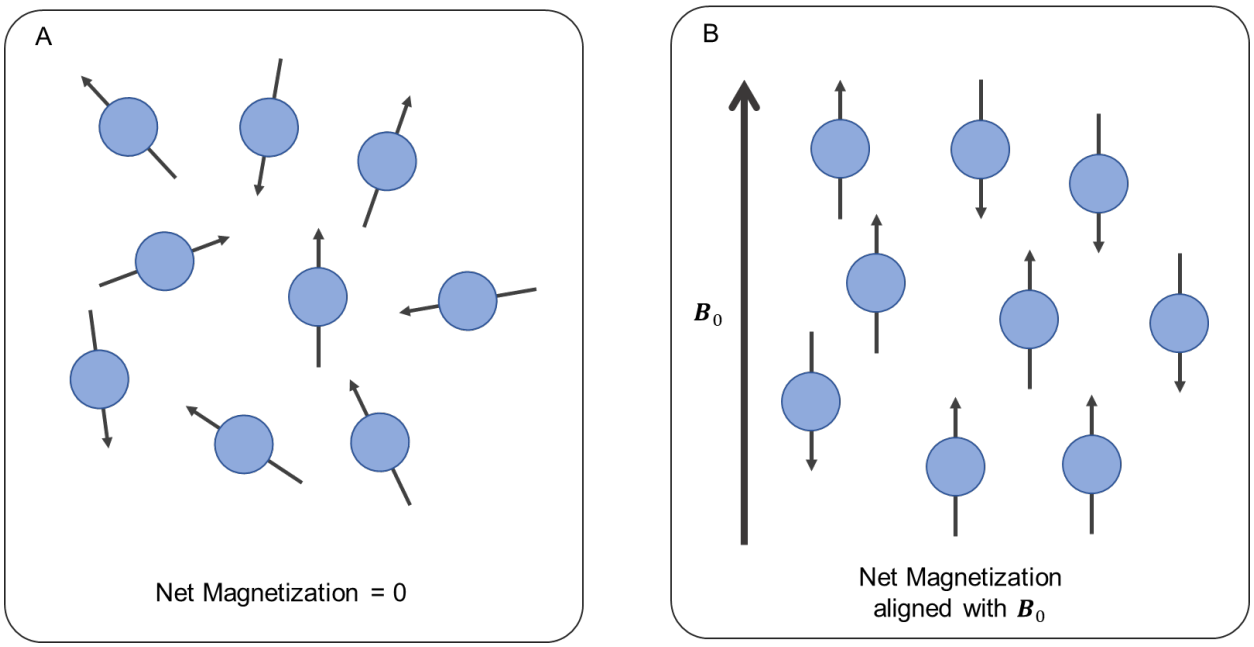

Figure 2.5 Magnetization of a system. (A) Without the presence of an external magnetic field, magnetic moments are randomly distributed, resulting in zero net magnetization. (B) When an external field is applied, spins align parallel to the direction of the applied field.

2.2.2.1 Nuclear Spins and Precession

Many biologically relevant atoms possess an intrinsic angular momentum due to having an odd number of protons and neutrons in the nucleus. The positive charge of protons in combination with nuclear spins result in these nuclei to act as magnets. A collection of identical nuclei, in their natural state, has zero net magnetization due to randomly oriented spins of individual nuclei. However, when placed in the strong magnetic field of an MRI, $\boldsymbol{B}_0$, the spins align in the direction of the applied field, generating a non-zero bulk magnetization with magnitude $\boldsymbol{M}_0$ in the longitudinal axis, as seen in **Figure 2.5**.

While in the applied magnetic field, nuclei undergo a phenomenon called precession, akin to the movement of a spinning top. The natural frequency at which nuclei precess is named the Larmor frequency, ω_0, and is given by:

$$\omega_0 = \gamma \boldsymbol{B}_0 \qquad (2.1)$$

Here, γ is the gyromagnetic ratio for each atomic species. ^{1}H is the species of interest in MR due to its natural abundance in the human body. In a 3T $\boldsymbol{B}_0$, typical of modern MR systems, ^{1}H protons precess at 127.74 *MHz*.

2.2.2.2 Excitation and Resonance

To induce a measurable signal, a second external magnetic field, $\boldsymbol{B}_1$, is applied. This process is denoted as a radiofrequency (RF) pulse and serves two main purposes by introducing energy into the system.

Under the $\boldsymbol{B}_0$ field, nuclei exist in either parallel or antiparallel states to the *z*-axis. Described by the Boltzmann distribution, the parallel direction (n^+) is in a lower energy state as it is aligned in the direction of the external field. With the introduction of $\boldsymbol{B}_1$, some nuclei gain sufficient energy to move to the antiparallel state (n^-), thus reducing $\boldsymbol{M}_0$ in the longitudinal direction. The Boltzmann distribution at equilibrium is described as such:

$$\frac{n^+}{n^-} = e^{\frac{\gamma\hbar\boldsymbol{B}_0}{k_B T}} \tag{2.2}$$

$\hbar$ is Plank's constant (1.054×10^{-34} *J s*), k_B is the Boltzmann constant (1.381×10^{-23} *J/K*), and T is temperature in Kelvins.

As nuclei precess along the longitudinal *z*-axis, magnetization in all other directions is net zero. The second effect of the applied $\boldsymbol{B}_1$ is the induction of all spins to precess coherently. This is performed using resonance, the phenomenon in which an oscillating force applied to a system will induce higher amplitudes. $\boldsymbol{B}_1$ is tuned to oscillate at the Larmor frequency of the nuclei of interest, thus energizing spins to precess in-phase with one another and generating a transverse magnetization in the *xy*-plane at the frequency

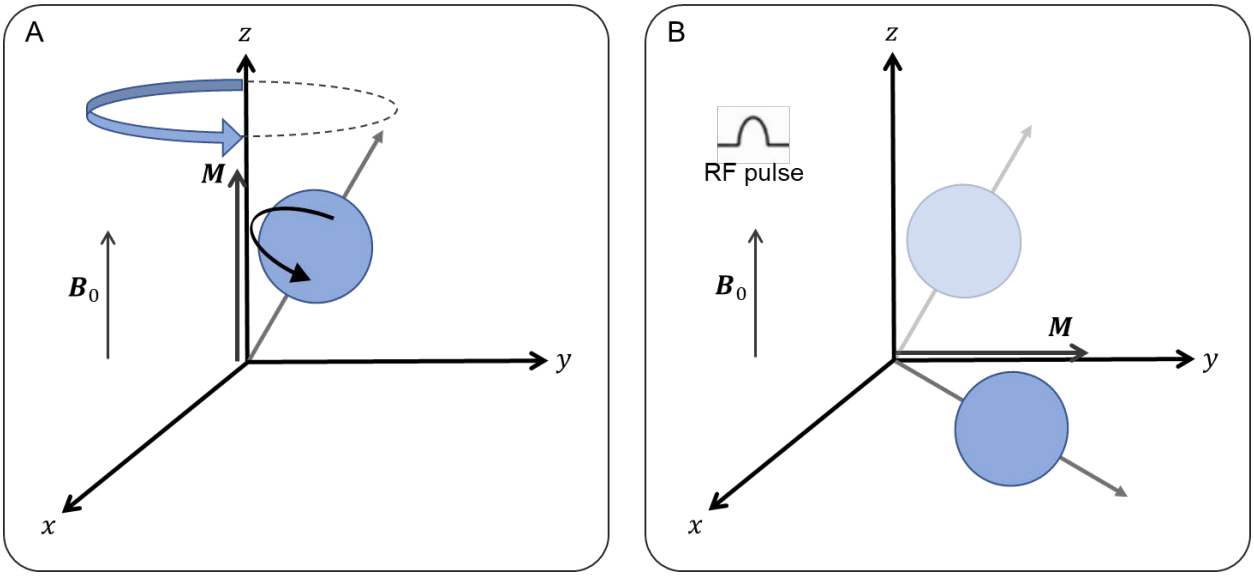

Figure 2.6 Precession and RF pulse excitation. (A) The nucleus precesses about the longitudinal *z*-axis and the collection of spins have a net magnetization in the direction of $\boldsymbol{B}_0$. (B) The application of an RF pulse tips the magnetization into the transverse *xy*-plane.

of precession. This net precession frequency is ultimately the source of signal induced in the RF receiver coil of the MRI system.

2.2.2.3 Relaxation and Image Contrast

The RF pulse introduces transverse magnetization and reduces longitudinal magnetization by adding energy into the system. Upon its removal, nuclei will naturally relax to their equilibrium states in the $\boldsymbol{B}_0$ field, restoring magnetization to $\boldsymbol{M}_0$ in the longitudinal axis and zero in the transverse plane. These are described as T_1 relaxation and T_2 relaxation (or decay), respectively. These time constants differ depending on the structure, biochemical composition, and environment of the objects being imaged which drives image contrast.

T_1 relaxation describes the process in which energy is exchanged between spins and their external environment through collisions, rotations, or electromagnetic interactions. Specifically, the term is the time required for the longitudinal component of

magnetization to reach $(1 - e^{-1})$, approximately 63%, of its maximum, $\boldsymbol{M}_0$ (**Figure 2.7**). Energy exchange in protons is heavily dictated by its composition. Free unbound water, for example, has relatively long T_1, approximately 4000 *ms*, due to its unrestricted nature being ineffective at energy transfer. Similarly, hydrogen protons bound to large macromolecules have low rates of motion and also possess long T_1, whereas protons in fatty acids have short T_1 relaxation as carbon bonds resonate near the Larmor frequency, promoting energy transfer.

While T_1 relaxation describes the regrowth of the longitudinal magnetization, T_2 relaxation, or more aptly named T_2 decay, describes the disappearance of transverse magnetization. This steady decay is caused by loss of phase coherence from internal inhomogeneity in the local environment. As spins precess together, their magnetic fields interact. Temporary random fluctuations of precession frequency in neighboring spins cause cumulative loss in phase, reducing overall transverse magnetization. Moreover, there is inhomogeneity in the magnetic field experienced by spins in different tissue composition and microenvironments. It follows that these small variations in magnetic field contribute to deviation from the Larmor frequency and subsequent loss of phase

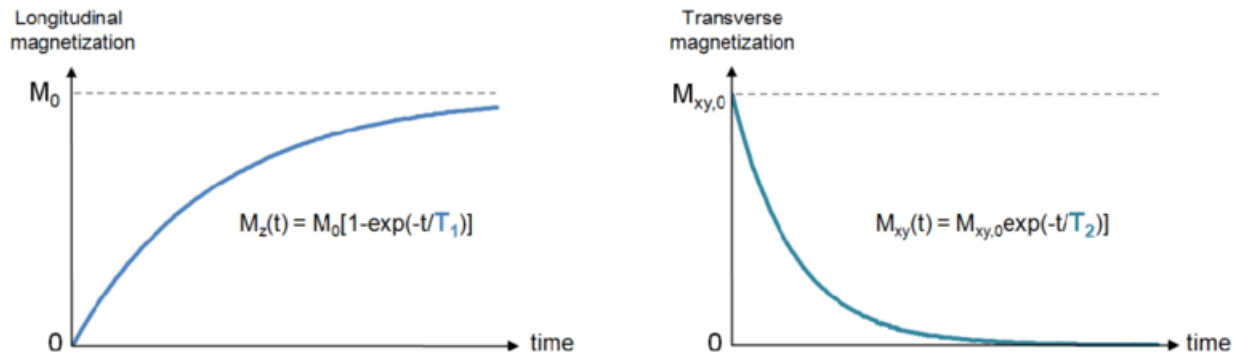

Figure 2.7 Relaxation rates of T_1 and T_2. Image source: Bosshard, *et al.* [57].

coherence. As with T_1 relaxation, T_2 follows an exponential trajectory and describes the time required to decay to e^{-1}, or approximately 37%, of $\boldsymbol{M}_0$.

Contrast in MRI is mainly governed by differences in T_1 and T_2 relaxation times of different tissues. To optimize for contrast in specific tissues, acquisition pulse sequences are designed to generate and measure signal at designated times. Three parameters that form the backbone of most image sequences include (1) repetition time (TR), the parameter between repeated RF pulses, (2) echo time (TE), the time at which electrical signal is measured, and (3) flip angle, the amount of rotation of the net magnetization during application of the RF pulse. By varying these parameters, the resultant images are weighted by intrinsic differences in T_1, T_2, or the minimization of their contribution, also known as proton density (PD) weighted imaging. This capability to design acquisition sequences optimal for contrast in targeted tissues gives MRI tremendous flexibility over other imaging modalities.

2.2.3 Standard MRI Acquisition and Analysis

In the clinic, MRI protocols are tailored with sequences most appropriate for the specific pathologic features to be evaluated while balancing total acquisition time, patient needs, and logistical constraints. A combination of T_1-, T_2-, and PD-weighted sequences fits most needs in MSK imaging. In spine MRI, a typical protocol consists of T_1-weighted and T_2-weighted fast spin-echo (FSE) sequences acquired in the sagittal and axial planes [58]. Fat suppressed sequences are often added to minimize fat signal and enhance contrast in other tissues. Additionally, a short tau inversion recovery (STIR) sequence may also be included for improved sensitivity of ligamentous and osseous lesions [59].

Protocols for whole joint assessment of the knee are similar and usually consists of imaging of all three orthogonal planes [60]. Each plane allows for delineation of crucial tissues. For example, the central femorotibial cartilage is best visualized in the coronal plane, while the patellar cartilage is optimally imaged axially. Offering the best single-plane assessment, the sagittal plane is particularly useful for delineation of the patellofemoral and femorotibial cartilage interfaces. Contrast is another consideration with respect to important tissue interfaces such as subchondral-cartilage, cartilage-synovial fluid, cartilage-meniscus, and cartilage-cartilage. These are generally best visualized using a combination of intermediate-weighted FSE, T_2-weighted FSE, and dual-echo steady-state (DESS) sequences. T_1-weighted FSE, while offering poor cartilage-fluid and cartilage-cartilage contrast, has excellent delineation of osteophytes and provides important information on bone marrow.

Grading of MSK abnormalities from imaging is almost wholly performed using semiquantitative scales. In knee OA, Whole Organ MRI Score (WORMS) [61] and MRI OA Knee Score (MOAKS) [62] are most commonly used [63]. Performing these assessments requires dividing the joint into subregions for feature-specific grading of OA, (*e.g.,* cartilage damage, meniscal tear or extrusion, bone marrow lesions, and osteophytes). Semiquantitative evaluation of spinal defects is less commonly used in the clinic. For example, the Pfirrmann grading system [64] is the most widely known classification for IVD degeneration, though its correlation with pain has shown to be inconsistent [65], [66] and is utilized mostly in research settings.

2.2.4 Advanced MRI Acquisition and Analysis

Standard image sequences used in the clinic are often generalizable to the assessment of many differential diagnoses of disease while maintaining reasonable scan times. The benchmarks of the abovementioned sequences are used ubiquitously due to their capacity for excellent depiction of gross morphology with the measurement of just a single TE per RF excitation. However, many advanced pulse sequences and analysis tools have been developed for proper quantification of tissue parameters and physiology beyond gross morphology. Compositional MRI, which measures cartilage matrix health, and high-resolution isotropic cube imaging are two technologies with relevance in future chapters that will be discussed here. Other acquisition techniques, like diffusion, perfusion, ultrashort TE or zero TE, or advancements in MRI reconstruction have been impactful in MSK imaging but are out of scope of this manuscript.

Detection of early biomarkers of disease has been a prominent focus in MSK imaging as it may allow for intervention while the condition is potentially reversible. As discussed earlier in this chapter, biochemical changes in tissue composition and organization precede many forms of degenerative disease. Compositional MRI describes a family of techniques with specific contrast weighting that have been empirically associated with distributions of contents in soft tissues. Sodium imaging and gagCEST, for example, correlate well with GAG content; however, acquisition times are lengthy and spatial resolution is poor [67]. dGEMRIC is another imaging technique to assess GAG content, though it requires administration of intravenous contrast. T_2 and $T_{1\rho}$ mapping have gained traction and are recommended approaches for compositional imaging. Parameterizing relaxation on a voxel-wise basis, these mapping techniques indirectly

measure water content of the ECM, orientation of collagen fibers, and GAG content in cartilage and have been validated in several studies [68]–[70]. However, a key limitation are long acquisition times. Proper estimation of relaxation times requires multiple data points per excitation pulse to fit to exponential relaxation curves, such as those shown in **Figure 2.7**. In addition, $T_{1\rho}$ acquisition requires the application of a secondary RF pulse parallel to the transverse plane to effectively lock the magnetization vector in the transverse plane [71]. Recent implementations have been able to achieve simultaneous acquisition of both T_2 and $T_{1\rho}$ mapping in a single sequence under 10 minutes in length. As acquisition capacities improve, we continue to see further push for the inclusion of these sequences in clinical imaging for improved detection of early degenerative disease.

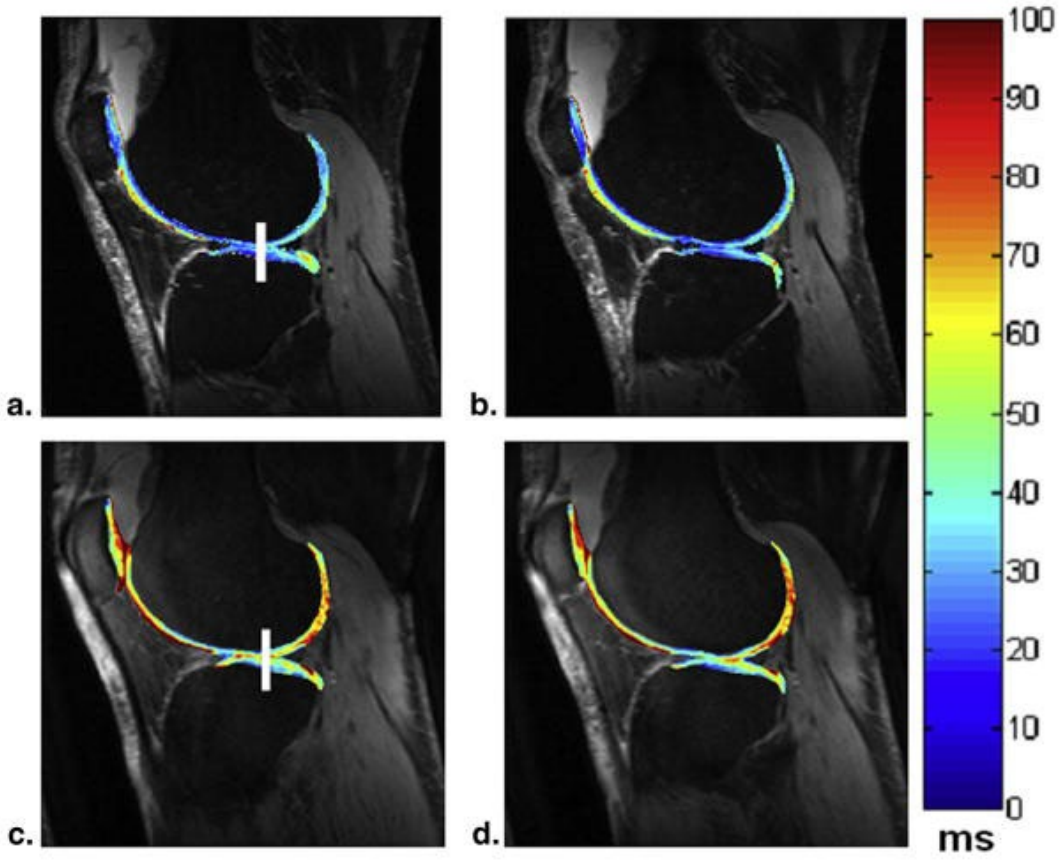

Figure 2.8 T_2 (a, b) and $T_{1\rho}$ (c, d) maps obtained from patient with OA. Image source: Wang L and Regatte RR [72].

Another challenge in effective quantification of MSK MRI is the concern of voxel size. Two-dimensional sequences used in the clinic are acquired slice-by-slice. While in-plane resolution is generally very high due to how signal is encoded in each RF excitation, through-plane resolution is sacrificed as acquisition time is proportional to the number of slices. Typical voxel sizes recommended for spine and knee imaging are 0.3125×0.3125×1.5 *mm* [58], [60]. Considering that the average thickness of femoral cartilage ranges between 1.95-2.59 *mm* [73] and vertebral endplates measure approximately 1.03 *mm* in thickness [74], abnormalities in the through-plane are easily missed. Partial volume effects are accentuated in the through-plane, as well, as poor resolution results in blurring and signal spillover between voxels. This limitation is somewhat mitigated by the inclusion of multiple 2D planar acquisitions, though evaluation of multiple planes in conjunction requires proper alignment, which is not always possible due to patient movement. Acquiring high-resolution images with isotropic voxels is certainly within reach but requires tradeoff with signal-to-noise ratio (SNR) due to fewer counts of hydrogen protons existing in smaller voxels. To counteract this loss in SNR, several strategies can be pursued [60]: (1) shorten TE to limit T_2 decay, (2) increase TR to increase T_1 recovery, (3) increase the number of excitations averaged, (4) use a system with stronger $\boldsymbol{B}_0$, or specialized coils. Vendors have implemented 3D volumetric FSE by leveraging a combination of these approaches with longer echo train lengths (ETL), ultrashort echo spacing, and parallel imaging. These techniques have demonstrated to have high clinical utility in MSK imaging and have become mainstay in many clinical routines in need of high-resolution, contiguous imaging [75], [76].

One key advantage to the techniques described above is the greater amount of total imaging information generated by these modern sequences. However, the repercussion is the greater cost in human labor to analyze them. Furthermore, quantification requires accurate segmentation of the tissues of interest. In the past, automatic and semiautomatic image processing techniques were capable of delineating tissue though were often troubled with abnormal morphology or imaging artifacts [77]–[79]. In the next section, we will lay the groundwork for DL-based image segmentation, which has much higher capacity for representation learning of image contents.

2.3 Machine Learning and Image Segmentation

Deep learning has rapidly demonstrated its potential in various applications, both within and outside of medicine. We will discuss the history behind the recent surge of DL, describe how deep neural networks have been adapted for semantic segmentation, and introduce strategies to overcome common challenges in generalizability.

2.3.1 History of Deep Learning and ImageNet

In 2009, ImageNet, the largest public image database at the time, was released with 12 million natural images across 22,000 categories to serve as a diverse training dataset for the advancement of computer vision research. One year later, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was born to invite the community to develop the best performing model in the classification of ImageNet [80]. During its lifetime, ILSVRC was the primary benchmark for vision tasks. However, progress was slow in its early years. It was not until 2012 when the first leap was made with AlexNet [81], scoring 9.8% ahead of its competitors in terms of top-5 classification error. Regarded as the

architecture that propelled DL to center spotlight, AlexNet was a multilayer convolutional neural network (CNN), incorporating many approaches that are still in-use today. In the years following, rapid improvements were made by other researchers in the field, modeled in various configurations of the CNN with deeper and deeper layers, thus spawning the era of deep learning.

2.3.1 Convolutional Neural Networks

Before we can discuss how the CNN impacted computer vision, we need to first understand the fundamentals of artificial neural networks (ANN). To illustrate this, we will describe the multilayer perceptron (MLP), a feedforward ANN, as seen in **Figure 2.9**.

In general, an ANN is a collection of nodes, where each node is defined by a mathematical transformation. A simple node may be a linear transformation in the form of $y = mx + b$, where x is the input, y is the output, and m and b are the node parameters or "weights". The multilayer aspect of the MLP describes the organization of the nodes in layers, with one input layer, one or more hidden layers, and one output layer. Following

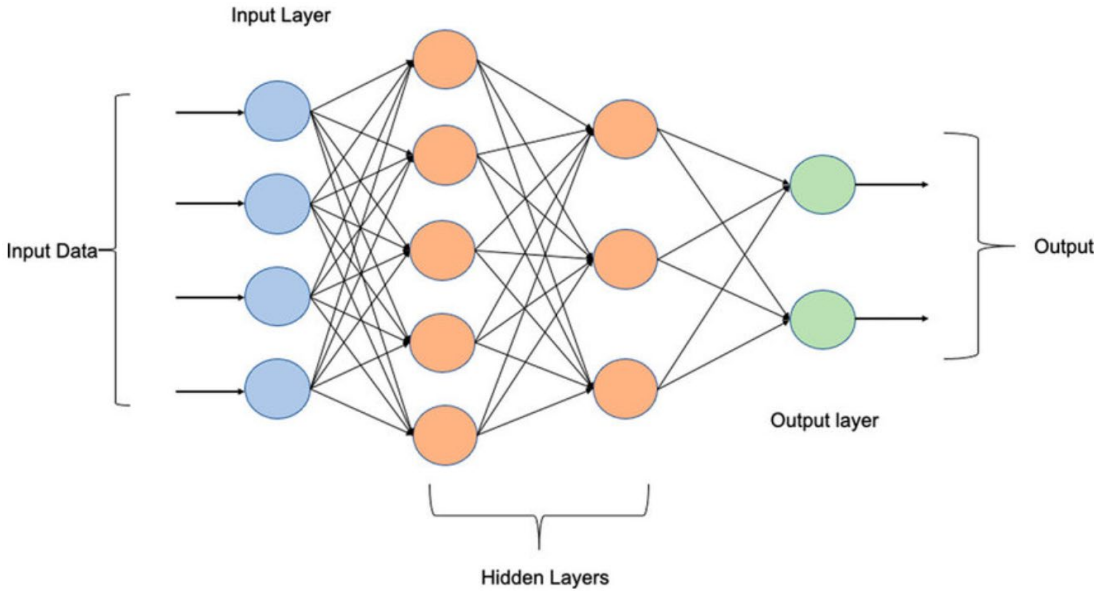

Figure 2.9 Schematic of multilayer perceptron. Image source: Afan HA, *et al.* [82].

each hidden layer, nonlinearity is applied with activation functions, such as a rectified linear unit (ReLU) [83]. The feedforward process describes the process in which inputs fed into the network are thereby transformed through the sequential hidden layers via the node weights to produce the output.

For the network to be tailored to a designated use case, the weights are adjusted in the training process. Input-output pairs are processed through the MLP in a feedforward fashion. Based on the amount of error between the “predicted” output and the expected true output, the weights in the neural network are adjusted in order to produce improved outputs in future iterations. The type of error is defined in each problem as the loss function. For example, in a regression task to predict an output on a continuous scale, the loss function can simply be formulated as mean square error. To determine the magnitude and direction of weight adjustment, this error is propagated backwards through the network, layer-by-layer. Leveraging the chain rule, a gradient is computed at each layer with respect to the current weights and a step is taken in the direction of the gradient to minimize the loss. This cycle of feedforward, error computation, and backpropagation is repeated with varying batches of data to iteratively optimize the model for minimal loss.

The traditional MLP architecture is well-defined for tabular data; however, it does not translate well to image data, in which each image pixel has no semantic meaning without context of its neighbors and suffers from multicollinearity. The convolutional layer addresses these issues and has been a building block in most vision architectures since AlexNet’s success in 2012. Based on the mathematical operation of a convolution, a convolutional layer consists of a set of filters or kernels that slides across an image while computing an elementwise dot product between the input and the filters at each spatial

position. In other words, the filters represent features to search for in the input, and the resultant feature maps represent spatial locations in which the filters were “activated” by the contents of the input. In a deep CNN with multiple convolutional layers, the initial layers capture low-level features (*e.g.,* edges, texture) while downstream layers respond to high-level features (*e.g.,* shapes, objects). The head of CNNs typically end with one or multiple fully connected layers, synonymous with an MLP, which utilize the features extracted from convolutional layers for task-specific purposes.

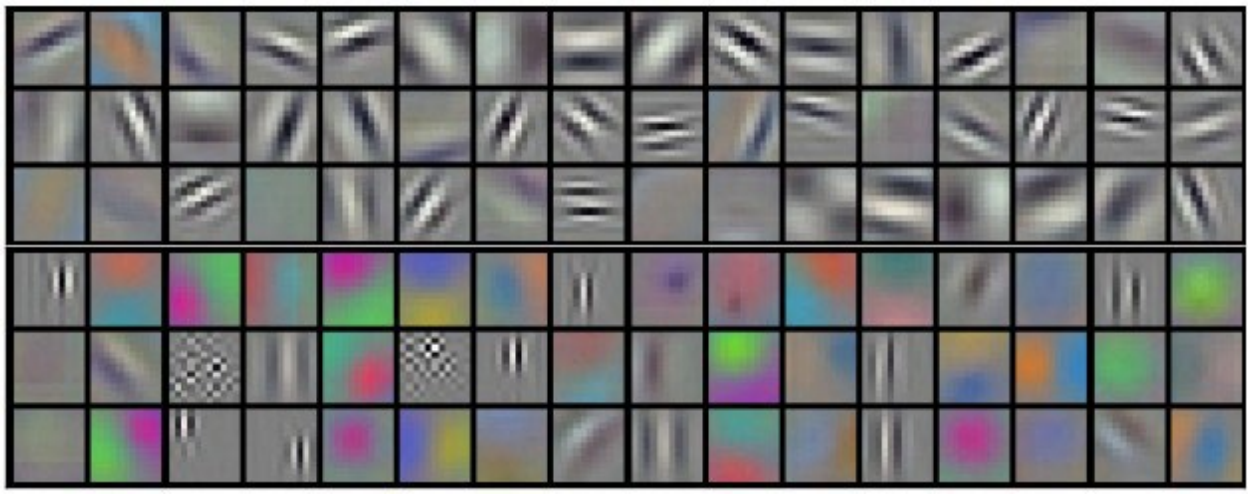

Figure 2.10 Kernels learned by AlexNet [81]. These filters represent image features learned by the convolutional layers. Most were frequency- and orientation-focused.

2.3.2 Encoder-decoder Architectures

The MLP and architectures described above transform an image to a single output label for image classification. In the semantic segmentation task, the objective is to attribute a label to every pixel in an image. Convolutional encoder-decoder architectures were developed to achieve this and have been the most prevalent form of ANNs for segmentation across the past decade, only recently being challenged by the Transformer architecture [84]. The encoder branch extracts features, condensing images to a hidden latent space representation. Then, the decoder branch unravels the hidden

representation to reconstruct an image-sized output. More technically, the contracting encoder consists of a series of convolutional layers followed by a pooling layer (*e.g.,* max pool). Meanwhile, the decoder mirrors the encoder, using transposed convolutional layers to expand the signal back to an image.

The U-Net [85] is the most prevailing version of the encoder-decoder architecture used for semantic segmentation [86]–[88]. One of its defining features (**Figure 2.11**) is the presence of skip connections, which feed the output of encoder levels to the respective mirrored level in the decoder to reduce overfitting, a topic discussed further below. Since the inception of the U-Net in 2015, many variants have been implemented

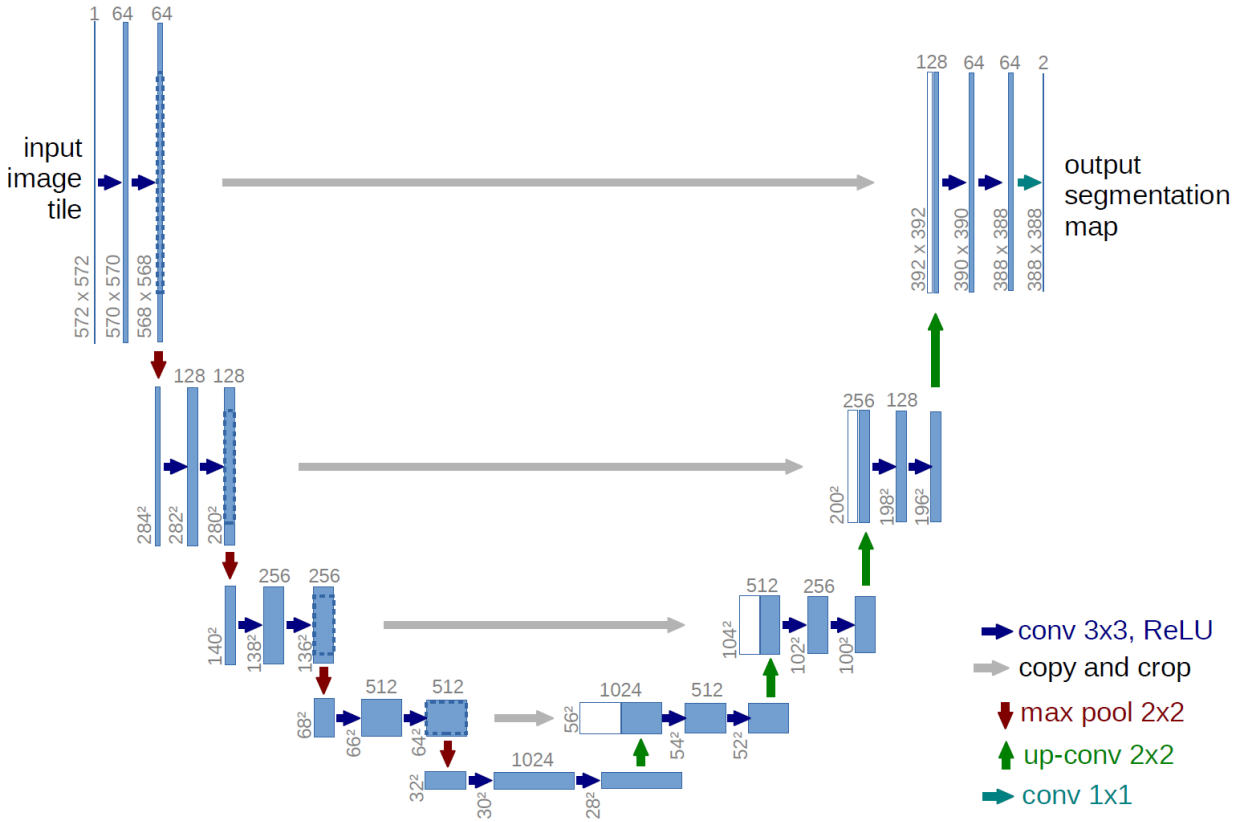

Figure 2.11 U-Net architecture. The contracting path (left) consists of repeated 3×3 convolutions followed by ReLU and max pooling for downsampling. The expansive path (right) consists of upsampling by a 2×2 and two 3×3 convolutions. Image source: Ronneberger, *et al.* [85].

with immediate success [87], [89]–[91]. The V-Net [90], heavily utilized in future chapters of this dissertation, was specialized for volumetric medical image segmentation. The key differences between the V-Net and its predecessor include a combination of 3D 5×5×5 and 2×2×2 convolutional layers, an element-wise sum after the series of convolutional layers, and an additional level of depth. Ultimately, regardless of the variant or architectural parameters, these fully convolutional encoder-decoders have transcended previous techniques for image segmentation.

2.3.3 The Challenge of Generalization

The primary goal of ML is to develop models generalizable for their designated use case. However, there are many obstacles that interfere with models' abilities to accurately predict outcomes from unseen data. These hurdles can be related to biased or limited training data, improper methodology, or insufficient model capacity. In the remainder of this chapter, we introduce several strategies to overcome these issues.

First, if not the most important, is the topic of data. In the traditional supervised learning paradigm, data is divided into a training set, and one or more evaluation sets. The training set, as its name implies, is used in the feedforward-backpropagation cycle described above to optimize and fit the model. The evaluation set(s) can be configured in different ways depending on experimental design. Most commonly, two evaluation sets are used. The validation set serves to assess model improvement during training, hyperparameter searches, and to inform model selection after training is complete. A test set is used to compute a final performance metric at the end of experimentation. Other validation techniques can be incorporated. For example, *k*-fold cross validation would require the division of the training data into multiple independent splits to better test model

generalizability. Ideally, samples would be randomly split between the training and evaluation sets. However, because it is imperative that each set is sampled from the same feature distribution, pseudorandom or stratified approaches may be more practical, particularly in cases with small sample sizes or heavy class imbalance.

Limited data is a common issue for model development in medicine. As modern CNNs have billions of learnable parameters, they are incredibly prone to overfitting to and memorizing the statistical noise of the training set. Regularization is general term to describe mathematical techniques that reduce overfitting. For instance, explicit regularization involves directly imposing a cost to model complexity. Lasso (L1) or ridge (L2) regularization can easily be implemented as additional terms in the loss function to penalize high weight values.

$$Loss\ (with\ L1\ regularization) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y} - Y\right)^2 + \lambda\sum_{i=1}^{N}|\theta_i| \tag{2.3}$$

$$Loss\ (with\ L2\ regularization) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y} - Y\right)^2 + \lambda\sum_{i=1}^{N}\theta_i^2 \tag{2.4}$$

The above formulations represent a generalized loss function with the inclusion of L1 or L2 regularization. Here, N is the number of samples, $\hat{Y}$ is the model prediction, Y is the true label, λ is a scaling hyperparameter, and θ represents the weight matrix.

Overfitting can also be tackled via stochastic training of multiple models. Model ensembling describes how predictions from multiple models can be combined for regularization, such as majority vote or weighted vote with posterior probability. Ensembling may be reasonable with abundant computational resources or with smaller models; however, several ML techniques exist to mimic the impact of model ensembling

without additional overhead. Similar to the effects of gradient boosting in decision trees, dropout has the effect of procedurally producing multiple weaker learners and combining their predictions. Dropout [92] refers to the probabilistic removal of nodes and their connections in the network architecture during training. The ensuing effect is the learning of sparser representations of the data. Dropout has shown to be effective in practice, synergizing well with other forms of regularization [92]–[94].

While the abovementioned regularization techniques target the training design and architecture, data can be regularized as well. The most common form of this is data augmentation, where training samples are randomly manipulated in an engineered fashion to generate additional examples that fit within the overall data distribution. Augmentation techniques in medical imaging are borrowed heavily from image processing to simulate variations in signal acquisition or morphological differences in tissue (**Figure 2.9**). Image synthesis techniques, such as with generative adversarial networks, are more advanced but have shown promise in improving model generalizability [95], [96].

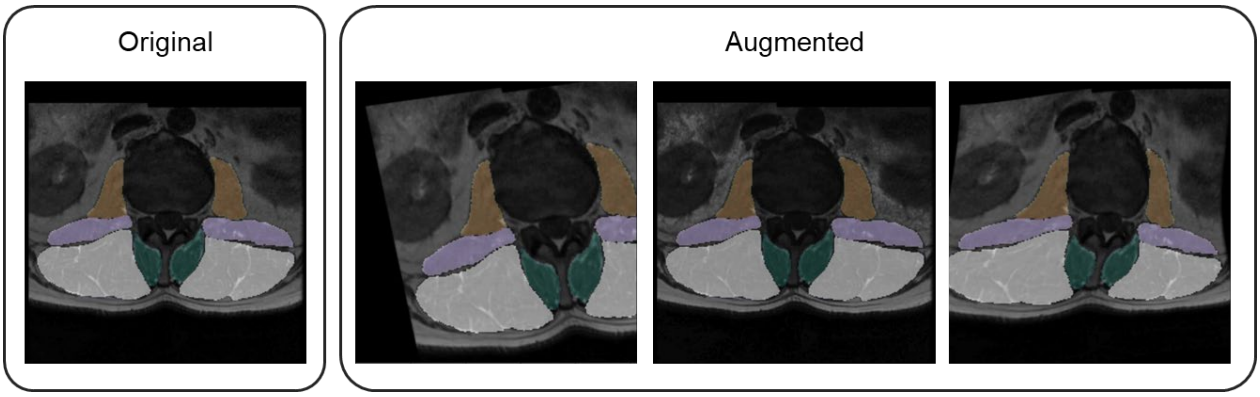

Figure 2.12 Example of data augmentation in lumbar spine MRI. The image and segmentation labels are manipulated in the conjunction to introduce plausible variance in the training set.

Transfer learning has demonstrated to be a powerful technique that allows for the application of past training to new model development. Alluding back to the deep CNNs trained on the millions of images in ImageNet, their weights can be finetuned for medical image analysis. Due to the hierarchical representation learning of deep CNNs, early layers which activate with low-level image features can be appropriated from the natural image domain. Adjacently, self-supervised learning [97], [98] describes the ML paradigm of using unlabeled data in a pretraining phase to build background knowledge prior to being finetuned in a specific downstream task. These techniques have dramatically closed the gap for vision tasks in challenging domains like medical imaging and bring us closer to overcoming the challenge of generalizability.

Chapter 3

Deep Learning for Multi-tissue Segmentation and Fully Automatic Personalized Biomechanical Models from BACPAC Clinical Lumbar Spine MRI

Robust automation of tissue segmentation has been a sought-after application of machine learning and image processing for decades. The recent introduction and advancement of convolutional encoder-decoder architectures like the U-Net [85] have propelled the reliability and accessibility of semantic segmentation models capable of delineating tissue from MRI. In this chapter, we describe the development and training of an ensemble of convolutional encoder-decoder networks for voxel-wise segmentation of tissues with established association with low back pain. This effort demonstrates the capacity of such networks to learn diverse representations of image data from limited sample sizes.

Furthermore, we showcase the application of these tools for developing musculoskeletal simulations of tissue loading.

The following material in this chapter is adapted and reproduced with full permission from the publisher. The original work is referenced as:

Hess M, Allaire B, Gao KT, Tibrewala R, Inamdar G, Bharadwaj U, Chin C, Pedoia V, Bouxsein M, Anderson D, Majumdar S. Deep Learning for Multi-Tissue Segmentation and Fully Automatic Personalized Biomechanical Models from BACPAC Clinical Lumbar Spine MRI. *Pain Med*. 2022 Oct 31:pnac142.

3.1 Introduction

Chronic lower back pain (cLBP) is a leading cause of disability in the United States [99] and is estimated to affect 540 million people worldwide [16]. Within the United States, cLBP is responsible for an estimated loss of 150 million workdays annually [100], [101] and is associated with an estimated annual cost of $100 to $200 billion [102], [103]. cLBP is one of the most common drivers of visits to a physician but, despite rapidly increasing treatment costs, patient outcomes have not improved substantially over time [104], [105]. cLBP is non-specific in 62.2% of cases [106], making patient-specific treatment interventions difficult to develop. The etiology of cLBP is multifactorial, including physical, psychological, environmental and socioeconomic factors [107]. Challenges in isolating the causes of pain can lead to overuse of imaging, opioids, and surgical treatment options.

Magnetic resonance imaging (MRI) and other medical imaging techniques are often used to design and monitor treatment strategies for patients with cLBP. Imaging alone is a weak predictor of pain presence and pain drivers due to a lack of consistent

associations between imaging studies and clinical symptoms [28], [28], [108]. There is debate in the literature regarding the relationships between different image-based morphological metrics and spine-related health indicators [28], [108]–[118], in part due to limited sample sizes and inconsistent measurement of imaging features across readers and institutions.

Segmentation of medical images is essential to unlocking scalable and reliable quantitative image-based markers of cLBP from spine morphology, but it can be prohibitively costly to perform. Manual segmentation is expensive and time intensive, requiring experienced readers between 15 minutes and several hours to annotate each exam, depending on number of slices and structures of interest. Further, despite the time and expertise investment required, annotations are subject to human error and bias across readers, making them difficult to generate and assess consistently in large-scale studies. Segmentation of the vertebral bodies, intervertebral discs, and paraspinous muscles in MRI can be valuable for diagnosing and characterizing spine degeneration and various pathologies related to cLBP, including stenosis, scoliosis and osteoporosis [119]. Estimation of internal tissue loading demands is a promising potential indicator for evaluating subject-specific risks for back pain and injury prevention [120]. Tissue loading cannot be directly measured, but biomechanical musculoskeletal models based on medical images can be used to provide valid estimates of tissue loading [121]. Segmentation of images of the spine is an essential step to creating subject-specific musculoskeletal models, which improve on generic models by incorporating measurements like spine curvature and muscle morphology [122]. If segmentation can be performed in a fast, low-cost manner, personalized biomechanical models of patients

could be automatically created as part of a clinical workflow. These models could then be used to evaluate measures of tissue loading as a component of back pain and to design patient-specific interventions.

Deep learning-based segmentation methods offer improvements in standardization and scalability when compared with human segmentation and can be leveraged to improve data collection in cLBP research. Combining computational methods with MR imaging presents opportunities for improved diagnosis with potential for multifactorial analysis, quantitative analysis, and increased sample sizes.

We propose a framework for automatic segmentation of vertebral bodies, intervertebral discs, and paraspinous muscles in clinically acquired sagittal and axial T_1-weighted MRIs of the lumbar spine and assess its efficacy as a substitute for manual segmentation in calculating intervertebral disc height, muscle cross sectional area and metrics of lumbar spine loading with subject-specific biomechanical models.

3.2 Materials and Methods

This study was approved by the University of California, San Francisco (UCSF) Institutional Review Board. A retrospective clinical dataset of 206 MR imaging exams with both axial and sagittal T_1-weighted acquisitions was aggregated as part of UCSF's Back Pain Consortium. All exams were randomly selected from clinical scans at UCSF between 2008 and 2018. Of that set, 24/27/45 (vertebral body [VB]/disc/muscle) exams were annotated with the respective anatomical structure. Cases with bone fractures, extensive implants, primary tumors, and wide-spread metastatic disease to the spine were excluded. All volumes included were drawn from clinical exams obtained on GE scanners and

Table 3.1 Lumbar spine image acquisition parameters

MR Imaging Sequence	Acquisition Parameters	Associated Network
T_1 Sagittal	TR = 500 – 735 *ms*, TE = 8.184 – 21.1 *ms*, ETL = 2 – 6, Acquisition Size = [0, 288, 224, 0] – [384, 0, 0, 224] Image Size = 512 × 512, Resolution = 0.4688 × 0.4688 *mm* – 0.5078 × 0.5078 *mm*, Slice Thickness = 3.0 – 4.0 *mm* Space Between Slices = 3.0 – 5.0 *mm*	Vertebral Body Segmentation Network Intervertebral Disc Segmentation Network
T_1 Axial	TR = 503 – 983 *ms*, TE = 7.672 – 17.376 *ms*, ETL = 3 – 6, Acquisition Size = [0, 256, 160, 0] – [320, 0, 0, 224] Image Size = 256 × 256 – 512x512, Resolution = 0.3125 × 0.3125 – 0.7031 × 0.7031 *mm*, Slice Thickness = 3.0 – 4.0 *mm* Space Between Slices = 3.0 – 5.0 *mm*	Muscle Segmentation Network

ETL – echo train length; *mm* – millimeter; *ms* – millisecond; TE – echo time; TR – repetition time;

followed the image acquisition parameters detailed in **Table 3.1**. Images were captured with subjects in head-first or feet-first supine position.

Manual vertebral body segmentations were performed on all slices in 24 exam volumes by one reader. Intervertebral disc segmentations were similarly annotated on all slices in 27 exams by two readers. Paraspinal muscle segmentations were annotated on one to three slices at each lumbar disc-level in 45 exams by one reader. Both readers were trained by a board-certified radiologist. All annotations were performed using MD.ai annotation software (**Figure 3.1**). When annotating the vertebral bodies and intervertebral discs, readers were instructed to segment all structures visible on every slice of a volume,

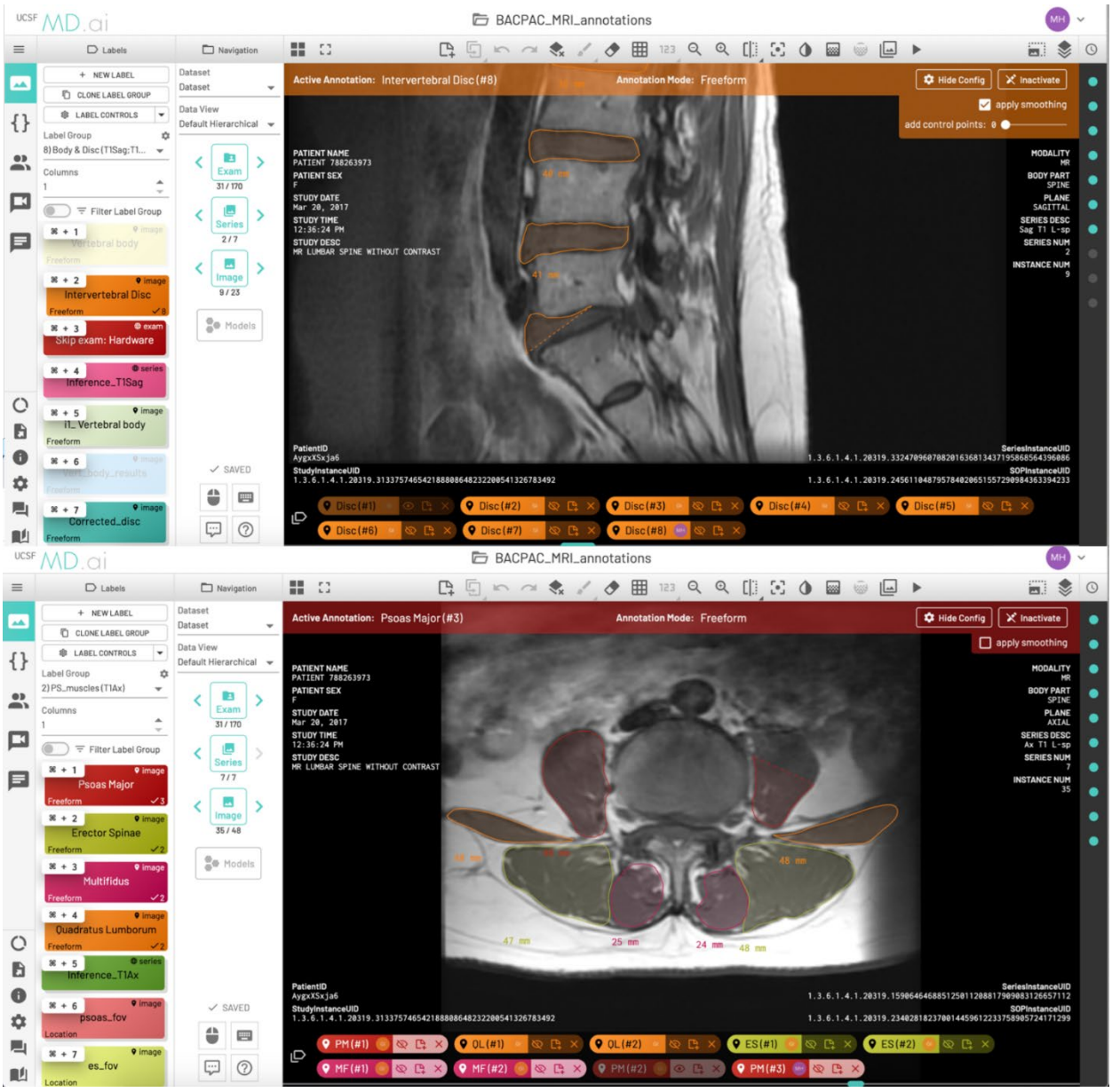

Figure 3.1 MD.ai annotation software. Each anatomical structure was annotated with a click and drag interface using the MD.ai annotation tool.

excluding (1) any bodies and discs that were not completely pictured in the field of view, and (2) structures inferior to the S1 vertebral body and L5-S1 intervertebral disc. All readers were trained to identify respective anatomy by a board-certified radiologist.

Separate 2D V-Nets [90] were trained to segment each anatomical structure on each sequence. The V-Net architecture and associated Dice-based loss function have

previously been demonstrated to have improved performance and convergence time on medical image segmentation tasks over other popular network architectures. A 2D instead of 3D approach was selected to maximize use of available data as the wide variation in number of annotated slices per exam would have required substantial slice-dimension coercion, which could impart bias through duplication or interpolation and loss of information through cropping. Because of slice thickness in clinical images is greater than the in-plane resolution, segmentation of these sequences was a good candidate for a 2-dimensional method. Two versions of each convolutional neural network were trained to segment the vertebral bodies, intervertebral discs, and paraspinous muscles respectively from T_1-weighted MRI volumes. Two version splitting was required to maximize both training performance and unbiased testing of biomechanical models; the standard split was constructed to demonstrate expected network performance under typical conditions, while the shared split was constructed to maximize samples in a hold-out test set to assess downstream performance without data leakage. The first version of each of the three models was computed with a standard random split of approximately 75% train, 15% validation, and 10% test to demonstrate optimal model performance on segmentation of each anatomical structure; we call this version the *standard split*. To accommodate limitations in data quantity, the networks were then retrained with new splits which reserved a shared set of 10 patients across models for testing to assess the performance of this segmentation pipeline on automatic morphological metric and biomechanical model generation; we call this version the *shared split*. All networks were trained according to the hyperparameters listed in **Table 3.2**.

Table 3.2 Training hyperparameters of spinal segmentation

Standard Split Networks			
	T_1 Sagittal Disc Segmentation	**T_1 Sagittal VB Segmentation**	**T_1 Axial Muscle Segmentation**
Network Architecture	2D V-Net	2D V-Net	2D V-Net
Input Image Size	512 × 512	512 × 512	256 × 256
Batch Size	8 slices	8 slices	32 slices
Learning Rate	1*e*-4	1*e*-4	1*e*-4
Dropout	0.2	0.05	0.2
Levels in Network	4	3	3
Optimizer	Adam	Adam	Adam
Loss Function	Weighted Dice Sigmoid	Dice Sigmoid	Weighted Dice Sigmoid
Iterations Until Convergence	14,700	12,000	13,500
Shared Split Networks			
	T_1 Sagittal Disc Segmentation	**T_1 Sagittal VB Segmentation**	**T_1 Axial Muscle Segmentation**
Network Architecture	2D V-Net	2D V-Net	2D V-Net
Input Image Size	512 × 512	512 × 512	256 × 256
Batch Size	8 slices	8 slices	8 slices
Learning Rate	1*e*-4	1*e*-4	1*e*-4
Dropout	0.2	0.05	0.2
Levels in Network	4	3	4
Optimizer	Adam	Adam	Adam
Loss Function	Weighted Dice Sigmoid	Dice Sigmoid	Weighted Dice Sigmoid
Iterations Until Convergence	17,500	11,900	12,000

Morphological metrics and biomechanical models were only evaluated for the networks in the shared split. Intervertebral disc, vertebral body, and paraspinous muscle segmentations were inferred for each patient in the shared hold-out test set of 10 exams, using the respective V-Net. Postprocessing was applied to all segmentations to smooth edges, fill holes, and isolate largest connected components. Intervertebral disc height (IVDH), muscle cross-sectional area (CSA), and 2D and 3D centroid positions for each anatomical structure in patient-based space were then extracted from the segmentations. These quantitative features were then leveraged as inputs to construct subject-specific biomechanical models of the lumbar spine to extract measures of compressive loading on the vertebral bodies.

To calculate IVDH, a 3D centroid was computed on each segmented disc to identify its most central slice; a minimum bounding rectangle was then constructed around the segmentation on the center slice to extract the shortest side length as a final height. Muscle CSA was constructed by calculating the sum of foreground pixels for each muscle, then multiplying by pixel spacing to yield area in square centimeters. Finally, a center of mass was computed in 3D to identify volume-wise centroids on each vertebral body and intervertebral disc, and in 2D to identify slice-wise centroids on each muscle (**Figure 3.2**). Each centroid point was then mapped to the patient-based coordinate system with an affine matrix transformation using the source exam's metadata, yielding results distance from the scanner reference point.

Subject-specific musculoskeletal models of the trunk were created with vertebral body centroids, muscle centroids, and muscle CSA in OpenSim [123]. Starting with an

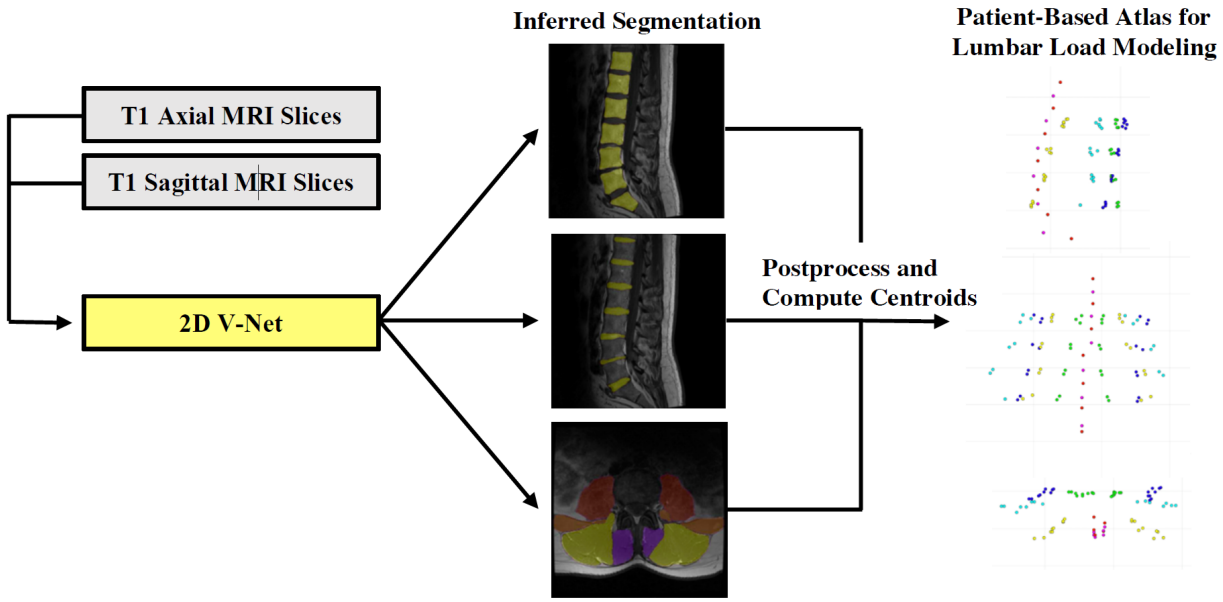

Figure 3.2 Visualization of centroid construction. T_1 axial and T_1 sagittal MRI slices were input into each respective V-Net to generate inferred segmentation masks of the vertebral bodies, intervertebral discs, and paraspinous muscles. After postprocessing, centers of mass were computed on each segmentation mask to calculate the position of volume-wise centroids for each vertebral body and intervertebral disc, and slice-wise centroids for each paraspinous muscle. These centroids were then converted to patient-based space, yielding a 3D atlas of the lumbar spine for further biomechanical modeling.

appropriate sex-specific base model, models were scaled to the subject's height and weight. Vertebral centroid locations and muscle morphology parameters were then incorporated into the model using custom MATLAB scripts to build the final model with subject-specific spine curvature and trunk muscle properties [122]. Lumbar spine compressive loading was evaluated using these models for a forward flexion activity (60 degrees of trunk flexion with 5 *kg* weights in each hand) at L1-L5 for each patient in the shared test set.

A volumetric Sørensen-Dice similarity coefficient [124] was calculated to assess the overall performance of each of the three networks. Each of the morphological metrics and biomechanical model outcomes were calculated on both the manual and inferred

segmentations for each patient, creating results that could be directly compared to assess the effectiveness of substituting manual for automatic annotation. Bland-Altman plots and Pearson R correlation coefficients were computed to assess the relationships between IVDH, CSA, and loading metrics generated from manually versus automatically segmented data.

Visual inspection for quality control after model training revealed several errors. Errors on the vertebral body annotations included segmentation of the S2 vertebral body or lower, exclusion of lateral slices in which bone was visible, and segmentation of both the vertebral body and bony pedicles in some of exams. One exam was annotated with paraspinous muscles on all slices. Major annotation errors were identified on one exam after training; the exam had been incompletely annotated and included less than half of the intervertebral discs and was consequently excluded from the hold-out test set on the shared split model.

3.3 Results

Volumetric Dice coefficients for each segmentation method are summarized in **Table 3.3**. The three segmentation networks with standard splits performed with Dice coefficients of 0.87 for the intervertebral disc network, 0.88 for the vertebral body network, and 0.81/0.95/0.84/0.92 (*mul*/*psoas*/*QL*/*ES*) for the muscle network on each respective test set. After re-splitting and retraining to construct set of the shared split networks, Dice coefficients decreased to 0.86 on disc, 0.76 on vertebral body, and 0.81/0.87/0.79/0.88 (*mul*/*psoas*/*QL*/*ES*) on muscle on the shared test set. Manual segmentations of the same exam demarcated by different annotators showed Dice coefficient similarities of 0.83 on

Table 3.3 Overall segmentation network performance on hold-out test set

Network		Intra-reader (n)	Inter-reader (n)	Standard (n)	Shared (n)
T_1 Sagittal Intervertebral Disc		0.88 ± 0.056 (3)	0.83 ± 0.039 (3)	0.87 ± 0.13 (2)	0.81 ± 0.047 (9)
T_1 Sagittal Vertebral Body		0.95 ± 0.013 (3)	0.93 ± 0.025 (3)	0.82 (1)	0.86 ± 0.033 (10)
T_1 Axial Paraspinous Muscle	Multifidus	0.88 ± 0.065 (3)	0.87 ± 0.049 (3)	0.81 ± 0.11 (2)	0.78 ± 0.082 (10)
	Psoas	0.94 ± 0.026 (3)	0.93 ± 0.028 (3)	0.95 ± 0.053 (2)	0.86 ± 0.078 (10)
	QL	0.79 ± 0.19 (3)	0.76 ± 0.21 (3)	0.84 ± 0.29 (2)	0.77 ± 0.078 (10)
	ES	0.92 ± 0.028 (3)	0.92 ± 0.029 (3)	0.92 ± 0.24 (2)	0.84 ± 0.092 (10)

Table 3.4 Stratified vertebral body network performance

Level	Shared Split ± 95% CI	Intra-Reader ± 95% CI	Inter-Reader ± 95% CI
T11	0.0243	0.947	0. 956
T12	0.888 ± 0.0218	0.968 ± 0.0113	0.957 ± 0.0252
L1	0.900 ± 0.0529	0.972 ± 0.0129	0.955 ± 0.0340
L2	0.888 ± 0.0642	0.971 ± 0.00449	0.939 ± 0.0341
L3	0.902 ± 0.0247	0.973 ± 0.00616	0.943 ± 0.00634
L4	0.863 ± 0.0753	0.972 ± 0.00542	0.951 ± 0.00957
L5	0.865 ± 0.0501	0.959 ± 0.0334	0.940 ± 0.0710
S1	0.684 ± 0.203	0.957 ± 0.0087	0.927 ± 0.162

Table 3.5 Stratified disc network performance

Level	Shared Split ± 95% CI	Intra-Reader ± 95% CI	Inter-Reader ± 95% CI
T11T12	0.54 ± 0.311	0.867	0.791
T12L1	0.844 ± 0.0567	0.886 ± 0.0165	0.864 ± 0.0471
L1L2	0.731 ± 0.213	0.848 ± 0.0850	0.799 ± 0.127
L2L3	0.813 ± 0.0586	0.868 ± 0.173	0.654 ± 0.804
L3L4	0.730 ± 0.220	0.908 ± 0.018	0.582 ± 1.25
L4L5	0.711 ± 0.216	0.871 ± 0.142	0.831 ± 0.116
L5S1	0.650 ± 0.200	0.899 ± 0.0470	0.554 ± 1.19

disc, 0.93 on vertebral body and 0.78 or greater on muscle. Manual segmentations of the same exam from one reader after a washout period of at least 3 weeks showed similar Dice coefficients of 0.88 on disc, 0.95 on vertebral body and 0.87 or greater on muscle. Example images of manual and automatic segmentations are shown in **Figure 3.3**. The vertebral body network performed with level-wise volumetric Dice coefficients greater than or equal to its overall performance of 0.76 on all levels (**Table 3.4**). Level-wise performance on the intervertebral disc network dropped below its overall performance of 0.86 on all disc-levels (**Table 3.5**). Results on the S2 vertebral body and S1S2 intervertebral disc are not reported, as annotators were instructed to exclude the S2 vertebral body and S1S2 intervertebral disc.

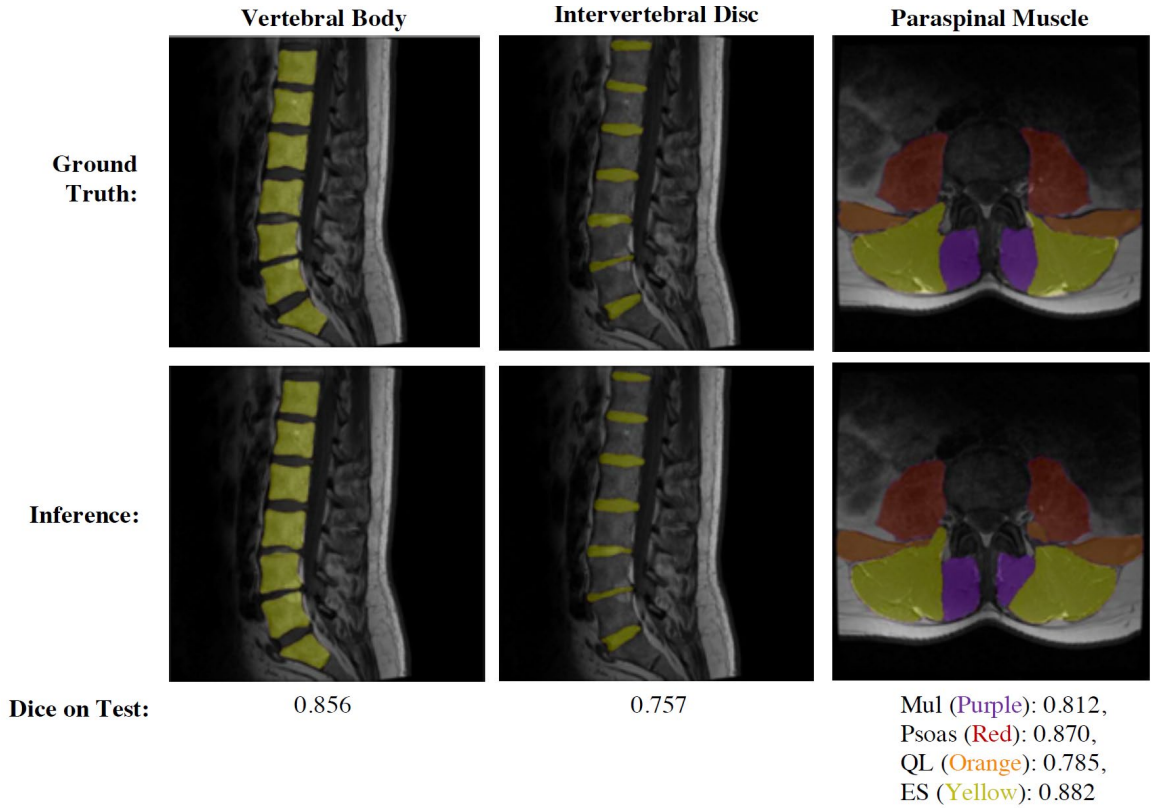

Figure 3.3 Visualization of segmentation results from each network. The first, second and third columns show examples of vertebral body, intervertebral disc, and paraspinal muscle segmentation results, respectively, along with a 3D Dice coefficient of each network's performance.

Table 3.6 Overall performance of automatically generated morphometrics

Biomarker	Pearson *R* (*p*-value)	Mean Absolute Error
Intervertebral Disc Height (mm)	0.846 (2.48×10^{-18})	1.45 ± 0.380
Muscle CSA – Multifidus (cm^2)	0.720 (3.50×10^{-63})	0.489 ± 0.122
Muscle CSA – Psoas (cm^2)	0.892 (2.36×10^{-135})	0.457 ± 0.152
Muscle CSA – Quadratus Lumborum	0.905 (1.34×10^{-145})	0.197 ± 0.0584
Muscle CSA – Erector Spinae	0.897 (4.08×10^{-139})	0.964 ± 0.208
Lumbar Loading (Newtons)	0.767 (8.29×10^{-10})	143. ± 50.3

Table 3.7 Stratified intervertebral disc height calculation performance (n = 9)

Level	Pearson *R* (*p*-value)	Mean Absolute Error (mm)
All Levels	0.846 (2.48×10^{-18})	1.45 ± 0.38
T11T12/ T12L1	0.951 (8.02×10^{-05})	1.05 ± 1.03
T12L1/L1L2	0.847 (0.00396)	1.23 ± 0.85
L1L2/L2L3	0.155 (0.69)	2.17 ± 1.81
L2L3/L3L4	0.696 (0.0372)	0.90 ± 0.52
L3L4/L4L5	0.795 (0.0105)	1.24 ± 1.02
L4L5/L5L6	0.562 (0.115)	2.24 ± 1.26
L5S1/L6S1	0.945 (0.000124)	1.32 ± 1.18

Table 3.8 Lumbar load performance (n = 9)

Level	Pearson *R* (*p*-value)	Mean Absolute Error (Newtons)
All Levels	0.767 (8.29×10^{-10})	143.26 ± 50.25
L1	0.699 (0.0361)	101.0 ± 104.45
L2	0.869 (0.00234)	136.66 ± 98.7
L3	0.795 (0.0105)	142.86 ± 146.32
L4	0.688 (0.0405)	163.64 ± 145.5
L5	0.547 (0.127)	172.14 ± 160.08

Pearson *R* correlation coefficients and mean absolute error were used to measure the relationship between morphological metrics calculated from manually versus automatically generated segmentations (**Table 3.6**). IVDH calculated from network generated segmentations versus manually generated segmentations showed a correlation coefficient of 0.26 and a mean absolute error of 1.98 mm between the two methods (**Table 3.7**). Muscle CSA from automatic segmentation was correlated with that of manual segmentations with coefficients 0.74/0.850.70/0.37 and mean absolute errors 1.21/1.36/0.82/3.55 cm^2 (*mul*/*psoas*/*QL*/*ES*). Centroid locations differed with mean absolute errors of 3.03/3.64/7.23/3.58 mm in Euclidean distance between ground truth and inference. Vertebral body compressive loading computed between inferred and manually generated input data was correlated with coefficients 0.93/0.90/0.78/0.68/0.52 (L1/L2/L3/L4/L5) as seen in **Table 3.8** and **Figure 3.4**. Bland-Altman plots for manually versus automatically generated intervertebral disc height and muscle CSA show correlation and agreement between the two methods (**Figure 3.5**).

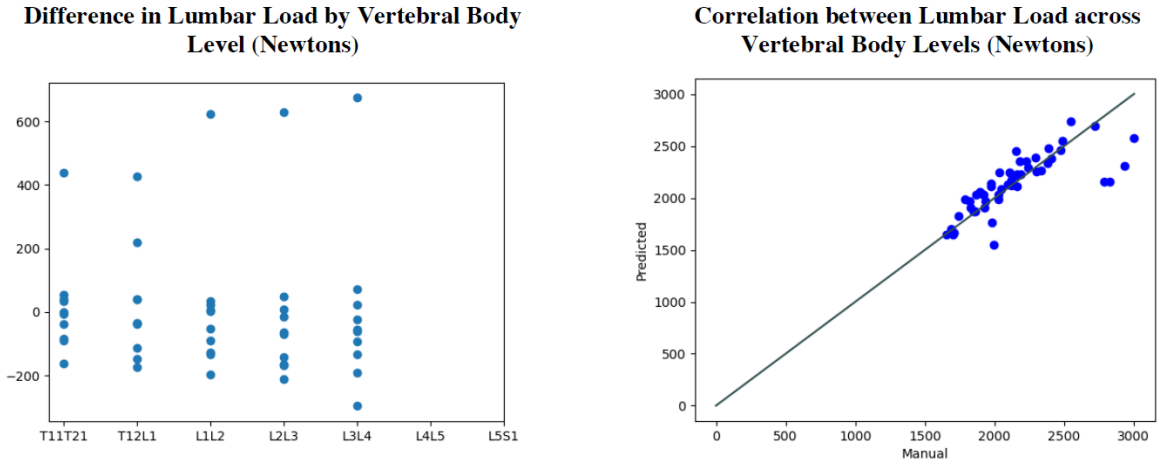

Figure 3.4 Stratified loading performance. Left, difference lumbar load in Newtons between automatic and manually generated results is plotted for each of the 9 exams evaluated from the hold-out test set at each lumbar level. Right, lumbar load calculated from manual segmentations is plotted against lumbar load predicted from automatic segmentations.

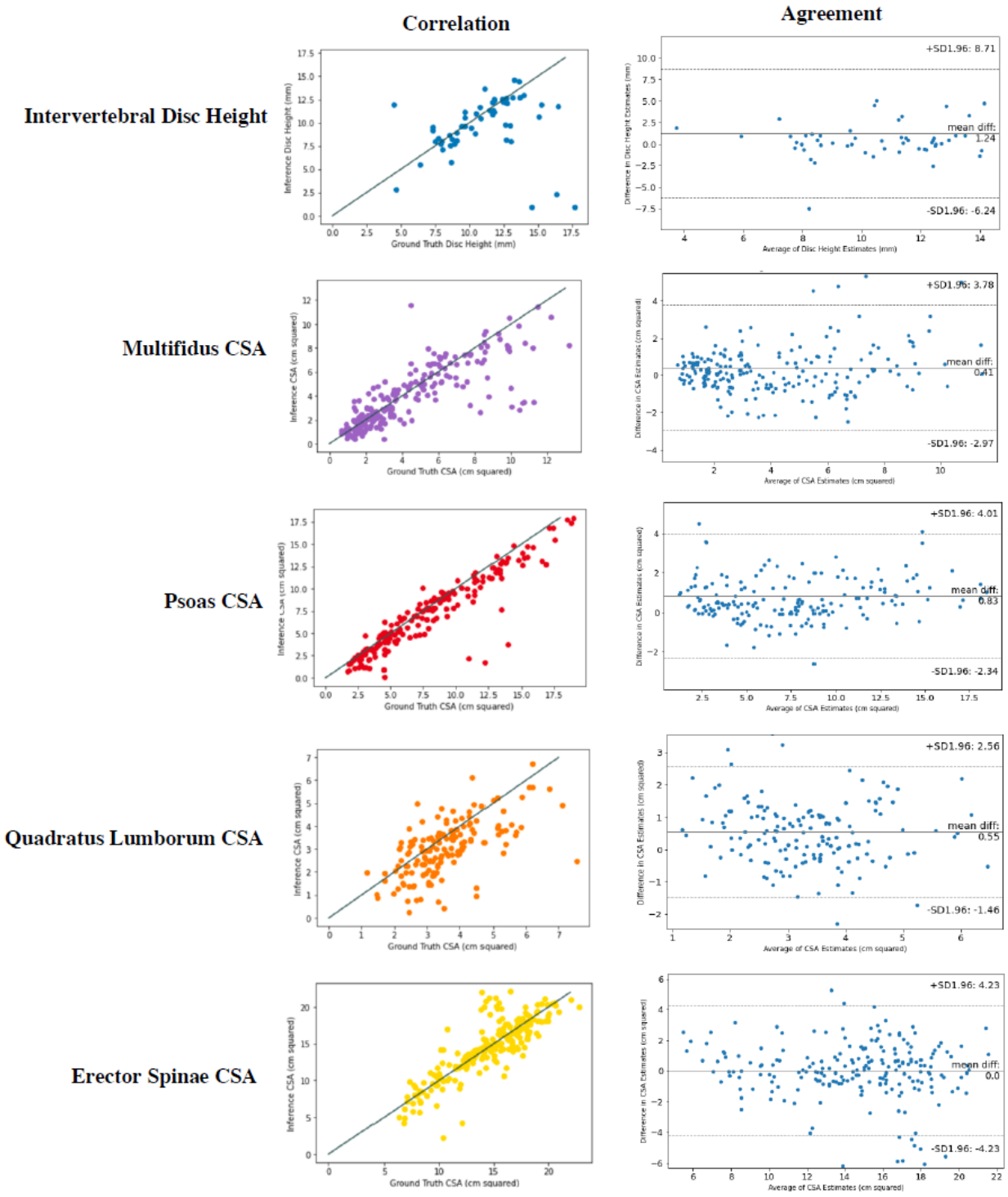

Figure 3.5 Correlation (left) and agreement (right) between manually and automatically generated segmentations for each biomarker. Correlation between disc height from manual versus inferred disc segmentations is displayed using a scatter plot, where the line *x=y* indicated in grey. Agreement is displayed using Bland-Altman plots for disc height on each disc.

3.4 Discussion

Previous studies have demonstrated the viability of automatic segmentation applied to medical images but have left notable gaps in the problem space in designing methods that integrate seamlessly into clinical workflow an create pathways to apply the technology at scale to biomechanical research. Demonstration of the feasibility of these networks to segment clinically acquired images is limited, instead requiring non-routine protocol or strict quality controls [119], [125]–[127]. Published networks do not apply this technology across planes and views to segment multiple anatomical structures or construct a fully subject-specific 3D atlas of the lumbar spine [128]. Additionally, published networks are specialized to downstream anomaly detection, not biomechanical modeling [129]. As a result, clinical translation of these methods is unfeasible on a large scale, as adding specialized imaging sequences to existing clinical protocol is costly and time intensive to institutions and patients. The generalizability of these methods to imaging studies routinely acquired in clinical settings is not yet proven.

We demonstrate the feasibility of substituting manual with automatic segmentation of the vertebral bodies, intervertebral discs and paraspinous muscles using networks trained on a small amount of clinical data. Overall segmentation network performance indicates that manual and automatic segmentation methods perform similarly, and morphological metric calculation can largely be outsourced to neural networks. While stratified performance results indicate value in human oversight of network performance and morphological metric generation, there are clear efficiency gains within an acceptable margin of error to be found by implementing fully automatic assistance when delineating vertebral bodies, intervertebral discs, and paraspinous muscle in T_1-weighted MRI. The

proposed segmentation pipeline and the associated quantitative feature generation methods have applications in both a clinical and a research context, as they will enable researchers to analyze larger datasets of potential biomarkers of cLBP and quickly provide those same features to clinicians to improve disease characterization and treatment in real time.

Performance for both the vertebral body and intervertebral disc segmentation networks is strongest on the central-most lumbar levels, which was consistent variation in training data and anatomical boundaries. Due to natural variation in patient anatomy, exams in model training exhibited a range in number of vertebral bodies and intervertebral discs in the lumbar spine. Annotators were instructed to exclude vertebral bodies that were not completely pictured in the field of view, as well as to exclude the S2 vertebral body or the S1S2 intervertebral disc but features of vertebral body and intervertebral disc boundaries are similar across the lumbar spine. This led to a pattern where trained networks inconsistently segmented the most superior and most inferior vertebral bodies and intervertebral discs in the field of view, as networks were arbitrarily penalized for correctly identifying these bodies in training. This phenomenon is reflected in the network's relatively poorer performance when segmenting the S2 vertebral body and T11T12 and S1S2 intervertebral discs (**Tables 3.4** and **3.5**). Segmentation of the multifidus, psoas, and erector spinae performance variance has no demonstrated correlation with slice level in most exams. Segmentation performance on the quadratus lumborum tends to drop on the most inferior and superior slices, consistent with anatomical expectation above L1 and below L5.

Error trends in automatic calculation of intervertebral disc height reflected trends in model error on the disc segmentation network, with statistical significance occurring only at the T12L1/L1L2 level. Results of all levels except L2 and L3 were inconclusive in the lumbar loading comparison. The decline in correlation from L1 to L5 may suggest greater sensitivity of lower lumbar loads to the subject-specific model inputs as models generate less reliable results on slices at the boundaries of anatomy. Automatic segmentation as an input may have tended to underestimate the height of each disc and load per vertebral body, but conclusions cannot be drawn given the small sample size.

We present a highly scalable, fully automatic framework to generate quantitative measures of spine morphology and subject-specific biomechanical models from lumbar spine MRI. Results generated by this pipeline are highly correlated and agree with those generated by human readers, without human-in-the-loop correction. This work indicates that computer-generated segmentations could successfully substitute for human-demarcated masks when quantifying metrics of lumbar spine morphology and biomechanical models to quantify tissue loading. These networks were trained on clinical exams with standard diagnostic sequences, suggesting strong generalizability with no extra costs associated with exam acquisition. A human-in-the-loop system to catch failures but improve time to acquire each segmentation could be of value to account for variation in performance with scan quality. A fully automatic, quantitative method for generating image-based features of spine morphology and validated estimates of tissue loading from clinically acquired MR exams combined with biomechanical modeling, like this one, would provide a scalable approach with which to evaluate drivers of cLBP across patients, institutions, and imaging archives without interrupting routine care.

Chapter 4

Automatic Detection and Voxel-wise Mapping of Lumbar Spine Modic Changes with Deep Learning

Structural tissue geometry is well-linked to conditions like OA and LBP with mechanically induced factors. Thus, it is crucial to understand the health and composition of the underlying tissue that give rise to adverse tissue shape. While compositional MRI techniques exist, these are often not available in boilerplate imaging protocols. This will be discussed with additional detail in **Chapter 5**. Rather, in the clinic, radiologists infer various states of tissue health using a combination of standard imaging sequences and semiquantitative grading protocols. Modic changes (MCs) are abnormalities along the vertebral endplate characterized by fibrovascular or fibrofatty tissue changes and are typically associated with degenerative disc disease. They have also been associated with a wide array of changes in bony composition depending on hypo- or hyperintense

appearances in T_1- and T_2-weighted MRI. However, due to technical factors such as the evolution of imaging systems and human variability between readers and studies, MCs suffer from a lack of conclusive evidence regarding their associations with LBP. In this chapter, we describe a quantitative approach, rooted in DL-based image segmentation, for developing a MC detection tool with utility in the clinic.

The following material in this chapter is adapted and reproduced with full permission from the publisher. The original work is referenced as:

Gao KT, Tibrewala R, Hess M, Bharadwaj UU, Inamdar G, Link TM, Chin CT, Pedoia V, Majumdar S. Automatic detection and voxel-wise mapping of lumbar spine Modic changes with deep learning. *JOR Spine*. 2022 Jun 8;5(2):e1204.

4.1 Introduction

Low back pain (LBP) is the leading cause of disability globally, accounting for 60.1 million disability-adjusted life-years in 2015 [2], [16]. While the nociceptive source in the vast majority of LBP cases cannot be identified [16], [17], there has been a growing collection of evidence showing that properties of vertebral endplates are closely linked to intervertebral disc degeneration and LBP [130]–[133]. Modic changes (MCs) are the most commonly used classification system for describing changes in endplate-adjacent vertebral bone marrow [134]. Despite its prevalence, the association of MCs with LBP is inconsistent [131], [135]–[137].

Hypothesized to cause LBP through structural and inflammatory changes in the bony structures of the spine [21], [138], [139], MCs are defined as signal variations seen in the combined assessment of T_1-weighted and T_2-weighted magnetic resonance

imaging (MRI) [134]. Bone marrow edema-like changes or fibrovascular changes appear distinctly hypointense on T_1-weighted images and hyperintense on T_2-weighted images (Modic type 1) [136], [140]. Meanwhile, conversion of red hematopoeitic bone marrow to yellow fatty marrow is hyperintense on T_1 MRI and iso- to hyperintense in fat saturated T_2 and non-fat saturated T_2 sequences, respectively (Modic type 2). And lastly, sclerotic bone appears hypointense in both sequences (Modic type 3).

Thus, the semiquantitative nature of the MC classification system is highly susceptible to variability in non-standardized imaging. Fields, *et al.*, detailed how evaluation of MCs are prone to inter-rater variability through a wide range of factors related to equipment and image acquisition parameters [140]. Magnetic field strength, in particular, has been shown to have significant effects on the prevalence of MCs, with type 2 changes being easily distinguishable in low-field MRI and type 1 changes visualized more easily in high-field MRI [141]. Pulse sequence design and parameters can also effectively influence image quality, signal-to-noise, fat suppression, and, importantly, tissue contrast. Due to a lack of systemic standardization in spine imaging, it is pivotal to adapt grading procedures with objective and quantitative methodologies.

Several quantitative approaches have been recently applied to the assessment of vertebral changes. Specialized pulse sequences, such as chemical shift encoding-based water-fat imaging [142], magnetic resonance spectroscopy [143], diffusion, and perfusion [144], can provide additional information on tissue composition. Post-acquisition, Wang, *et al.*, extracted morphological and signal intensity-based metrics from contours of MCs, reporting improved inter- and intra-rater agreement as compared to unassisted MC

classification [145]. However, a limitation with these approaches is the need for manual demarcation of MCs, which may be labor-intensive.

Data-driven strategies to address these drawbacks have emerged from the recent surge of development in deep learning (DL) and convolutional neural networks. Notable applications to spinal imaging analysis include automated segmentation of spinal structures [146]–[148], detection of spinal anomalies [149]–[151], and predictive modeling of spinal surgery outcomes [152], [153]. Automated endplate assessments have seen relative success, as well. Jamaludin, *et al.*, has shown that endplate defects can be detected from MRI using convolutional neural networks with approximately 83.7 and 86.9% accuracy in their test set for upper and lower endplates, respectively [154]. While these efforts automate spinal analysis to near human-performance, there remain opportunities to translate such models into clinical utility.

The adoption of a DL model into widespread use to address inconsistencies of the assessment and reporting of MCs hinges on its interpretability. Our study aims to (1) develop a DL-based automatic contouring method to identify MCs in vertebral bodies, (2) classify these changes as Modic types 1, 2 or 3 on a voxel-wise level, thereby providing granular, quantitative information about the vertebral bodies as a *Modic map*, and (3) use the automatic detection as an aid to radiologists, improve agreement, and pave the way for more consistent evaluations of the relationship between MCs and LBP.

4.2 Materials and Methods

This retrospective, single-center study was approved by the local Institutional Review Board and the informed consent requirement was waived.

Table 4.1 Acquisition parameters of curated clinical lumbar spine MRI exams

	T_1-weighted	**T_2-weighted**
Field Strength (*T*)	1.5, 3.0	1.5, 3.0
Matrix	256×256-512×512	256×256-512×512
Field-of-view (*cm*)	24.0-37.0	24.0-37.0
Slice thickness (*mm*)	3.0-4.0	3.0-4.0
Pixel bandwidth (*Hz*)	88.8-250.0	81.4-325.5
Repetition time (*ms*)	377-975	2430-6307
Echo time (*ms*)	6.8-31.8	26.1-107.8
Flip angle (°)	90-180	90-160

T – Tesla; *cm* – centimeter; *mm* – millimeter; *Hz* – Hertz; *ms* – millisecond; ° - degrees

4.2.1 Dataset and Annotations

Seventy-five exams with the following inclusion and exclusion criteria were sampled at random from lumbar spine MRIs acquired between 2008 and 2019 at our institution. *Inclusion:* patients aged 19 years or older presenting with acute-to-chronic LBP, radiculopathy, and other symptoms of the lumbar spine including numbness, tingling, weakness, dysesthesia, and tightness. *Exclusion:* (1) vertebral fractures, (2) post-operative changes, (3) extensive hardware, (4) primary tumors, (5) metastatic spinal disease, (6) infection, and (7) transitional anatomy. Imaging was performed on GE Signa HDxt 1.5T and GE Discovery MR750 3.0T (GE Healthcare, Milwaukee, WI) with acquisition details of the relevant T_1-weighted sagittal and T_2-weighted sagittal sequences provided in **Table 4.1**. All images were deidentified for this study.

To serve as ground truth for the DL components, vertebral bodies with visible MCs were segmented for these changes (Type 1, 2 and 3) by a board-certified neuroradiologist with 25 years of experience and a musculoskeletal junior radiologist with 3 years of

experience after initial adjudication for calibration on 15 exams not included in the study cohort. To promote further standardization between grading assessments, MCs with diameter less than 5 *mm* were excluded and mixed MCs were annotated as the predominant type. All manual annotations were performed using the medical imaging platform, MD.ai (MD.ai, New York, NY).

4.2.2 Image Analysis

This Modic mapping scheme consists of three stages, as depicted in **Figure 4.1**: (1) segmentation and localization of the vertebral bodies, (2) binary detection and segmentation of signal variabilities characteristic of MCs, and (3) voxel-wise classification of the detected regions to classify Modic type.

Image Alignment

As MCs are characterized by local signal variations in both T_1- and T_2-weighted images, these images were aligned with image position coordinates prior to processing. The rigid alignment was performed by first matching positions of each sagittal slice of the T_2-weighted images to the T_1-weighted images in the frontal axis. Then, T_2-weighted slices

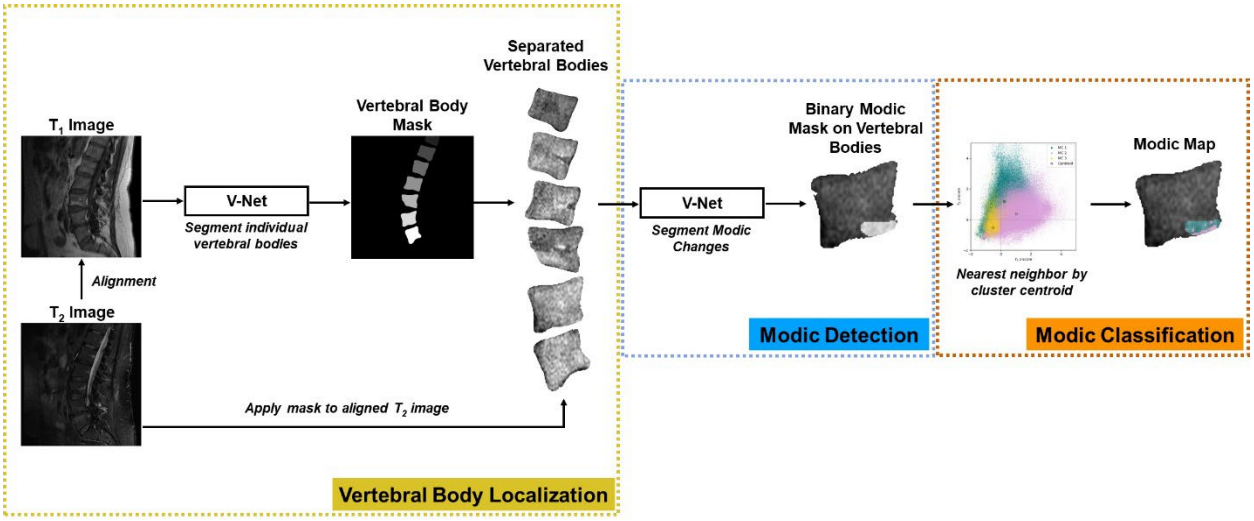

Figure 4.1 Schematic of the full Modic mapping approach.

were rotated, translated, and scaled to the dimensions of their corresponding T_1 counterpart. Finally, each slice was similarly translated and scaled to harmonize in-plane resolution using bicubic interpolation.

Vertebral Body Localization

Our first goal was to isolate vertebral bodies to fixate on image features pertaining to the vertebral body and endplates. To achieve this, we developed and trained a preliminary V-Net convolutional neural network [90] for semantic segmentation. A research associate manually segmented vertebral bodies from T_1-weighted images in a subset of 40 exams. These MRIs were randomly split into training (*n* = 20), validation (*n* = 17), and test (*n* = 3) sets and then separated into 2D slices. The V-Net was trained on a single NVIDIA TITAN X GPU using Tensorflow v1.14 with the following hyperparameters: batch size = 3; optimizer = Adam; learning rate = 1*e*-4; loss function = Dice (4.1); dropout rate = 0.8. Post-training, the performance of the segmentation model was assessed using the Dice coefficient overlap between the manual and predicted segmentations. To evaluate inter-rater variability, a second research associate manually segmented vertebral bodies from a subset of 5 exams.

$$Dice\ Loss = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{4.1}$$

where N is the total number of voxels, $p_i \in P$ represents voxel values of the prediction, and $g_i \in G$ represents voxel values of the ground truth.

We utilized this model to segment vertebral bodies of the 75 lumbar spine MRI exams in the dataset. The individual vertebral bodies in the inferred masks were identified using 3D connected component labeling, in which segmented masks joined within a 6-

connected neighborhood were given a unique label. The masked vertebral body masks were then zero-padded to a standardized size of 100×100.

Modic Detection and Segmentation

MC detection was achieved using a second segmentation neural network which utilized these localized vertebral bodies and the radiologist-annotated MCs. In each exam, we used *z*-score standardization to convert each voxel to the number of standard deviations from the mean signal intensity in the segmented vertebral bodies. Next, the 100×100 vertebral body masks were applied to the T_1-weighted and aligned T_2-weighted images and these images were stacked, producing input images of dimensions 100×100×2. Binary radiologist-annotated MC segmentations (presence versus absence of MCs) were similarly masked. The 75 exams, consisting of 1,872 vertebral body image-Modic segmentation pairs, were randomly split into training (*n* = 50), validation (*n* = 15), and test (*n* = 10) sets. **Figure 4.2** portrays the demographic distribution of the data splits.

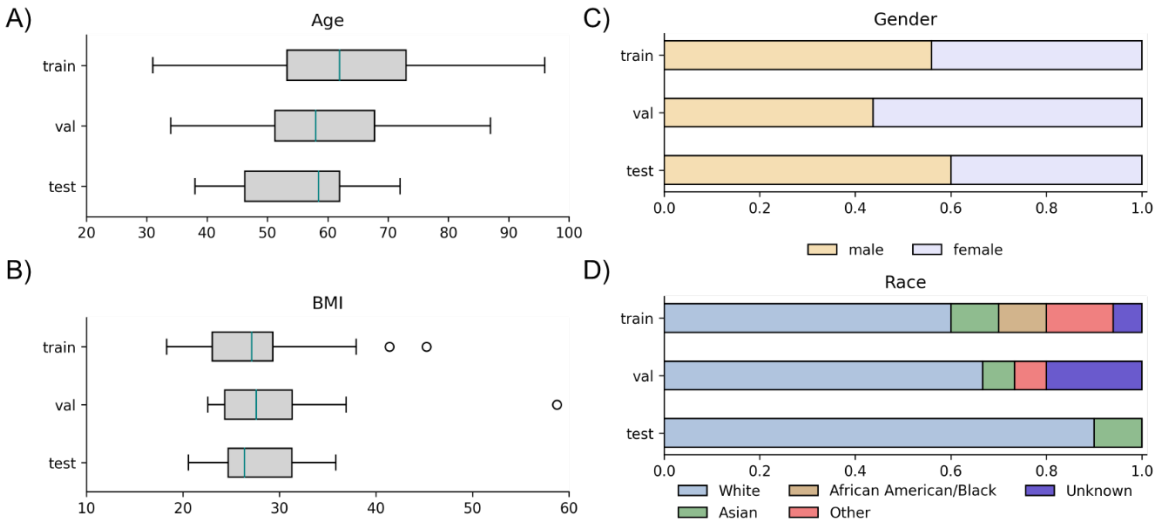

Figure 4.2 Distribution of subject demographics of the 75 lumbar spine MRI exams after randomly splitting into training (*n* = 50), validation (*n* = 15), and test sets (*n* = 10).

We developed and modified the 2D V-Net for MC segmentation. The network consists of two branches, each with 4 levels. The encoder branch is responsible for compressing the input to an abstract latent space of representative features. At each level, convolutional layers (1, 2, 3, and 3 layers in the respective levels) extract features with 32 kernels of size 5×5 and stride 1 followed by downsampling with a 2×2 kernel with stride 2. The subsequent decoder branch deconvolves the latent space back to the input's original dimension and passes the array through a combined cross-entropy and Dice loss layer with sigmoid activation to ultimately produce probabilistic segmentation masks for MCs. Hyperparameters for training include: batch size = 128; optimizer = Adam; learning rate = 1*e*-4; loss function = weighted cross entropy and Dice (4.2); loss weights = 20:1 (foreground:background); dropout = 0.2. Training was deemed complete after a designated 15 validation cycles without improvement (500 iterations per cycle).

$$Combined\ Loss = \lambda(Cross\ Entropy\ Loss) + \ Dice\ Loss \tag{4.2}$$

where λ is a weighting coefficient set to 0.1, and

$$Cross\ Entropy\ Loss = -\sum_{i=1}^{N} g_i \log(p_i) + (1 - g_i) \log(1 - p_i) \tag{4.3}$$

Voxel-wise Modic Change Mapping

With a trained model for Modic segmentation, we then utilized a nearest-neighbor algorithm to classify each voxel in the detected MCs into one of three types. Again, we utilized the training set; each voxel in the regions annotated by the radiologist was characterized by its T_1 *z*-score and T_2 *z*-score and then grouped into the appropriate MC group. The centroid of the [T_1 *z*-score, T_2 *z*-score] clusters were computed. To classify the test set and exams in inference, each voxel in detected MCs were similarly

characterized by [T_1 *z*-score, T_2 *z*-score] then categorized by the nearest cluster centroid neighbor. This ultimately produced voxel-wise Modic maps.

4.2.3 Statistical Analysis

We created a rule-based algorithm that produces binary labels of each MC in upper and lower vertebral bodies to assess the effectiveness of this scheme as compared to human performance and past works. Upper and lower sections were approximated by finding the convex hull of the vertebral body mask and bisecting them along the long axis. Thus, each bisection was described with three binary labels, representing the presence or absence of voxels characteristic of Modic types 1, 2, and 3, respectively. Sensitivity, specificity, and Cohen's kappa score (κ) were computed to evaluate the overall Modic detection performance, and the subsequent classification.

AI-assisted Experiment

A second dataset (n = 20) was curated to explore the effect of inter-rater agreement of Modic grading with the aid of this Modic mapping pipeline. A senior neuroradiologist (over 25 years of experience), a senior musculoskeletal radiologist (over 25 years of experience), and a junior radiologist in-training (3 years of experience) graded these exams independently. Inter-rater reliability was assessed using Cohen's kappa coefficient. After a four-week washout period, the musculoskeletal radiologist and junior radiologist re-graded the same dataset, with the aid of Modic maps generated from our developed pipeline. Agreement was reassessed to measure differences with the initial trial using Cohen's kappa score and the McNemar's test, with the neuroradiologist established as the baseline. The experimental setup is summarized in **Figure 4.3**.

4.3 Results

Vertebral Body Localization

Training the vertebral body segmentation network completed in approximately 10 hours with 20,000 iterations. Evaluated with the unseen test set, the model achieved 0.882 ± 0.018 Dice overlap with the ground truth segmentations. This performance is comparable to the inter-rater Dice overlap between two research associates, which was reported as 0.927 ± 0.011.

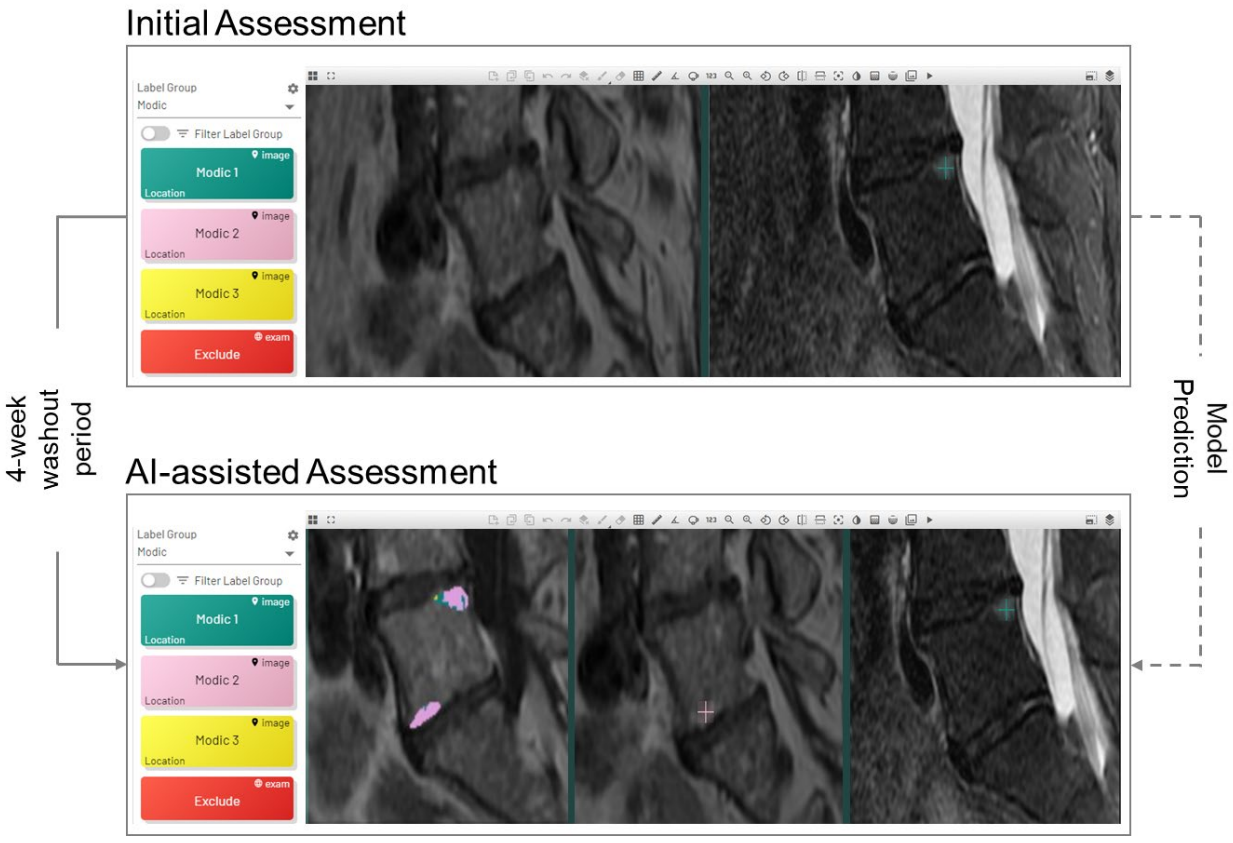

Figure 4.3 Experimental setup of the AI-assisted assessments in the labeling platform, MD.ai. Three readers graded an independently curated dataset (*n* = 20). Using the trained Modic mapping schema, predictions for MCs were generated in the same dataset, and after a 4-week washout period, readers 2 and 3 re-graded these exams with the assistance of the model predictions.

Post hoc analysis of vertebral body segmentation was performed (**Figure 4.4**). The mean volumetric error of the model prediction was 0.28 cm^3 per vertebral body, approximately 1.1% of the average vertebral volume. Manually segmented ground truth and model predictions were well correlated with an *R*-value of 0.94 and *p*-value < 0.001 using Pearson correlation.

Modic Detection and Segmentation

The Modic detection model, after training for 11,500 iterations, successfully identified the presence or absence of changes in 85.7% of samples in the unseen test set. Sensitivity and specificity of the model were computed and summarized in **Table 4.2**, resulting in 0.71 (±0.072) and 0.95 (±0.022), respectively. Cohen's kappa score was similarly computed against the radiologist-annotated ground truth as 0.63, interpreted as substantial agreement.

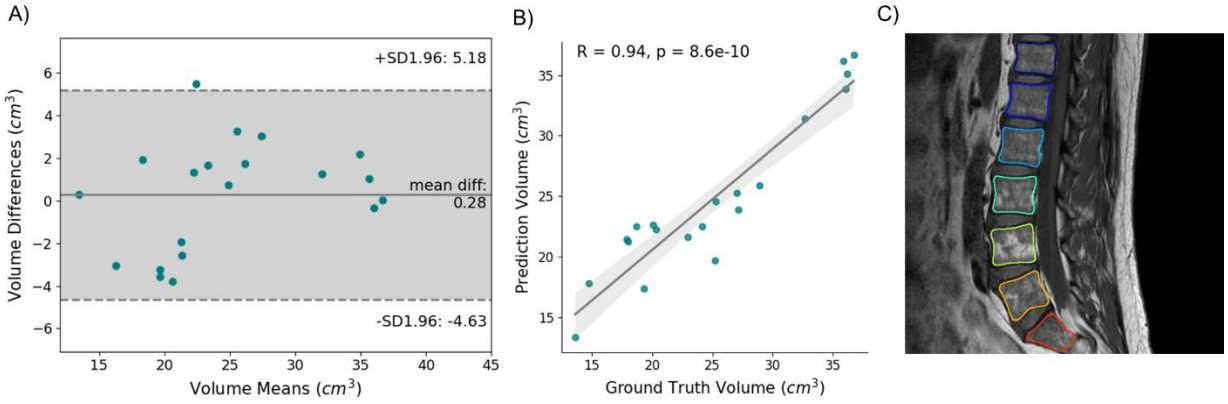

Figure 4.4 Post hoc analysis of vertebral body segmentation of the test set. (A) Bland-Altman plot indicate the average difference in vertebral body volume between model prediction and ground truth was 0.28 cm^3. The grey areas portray the 95% confidence intervals. (B) The correlation plot of vertebral body volume has an intercept of 14.8 cm^3, demonstrating a measurement bias, and *R*-value of 0.94. (C) Representative example of vertebral body segmentation contours on T_1-weighted image.

Table 4.2 Performance of the full Modic pipeline on the unseen test set

	Sensitivity (95% CI)	Specificity (95% CI)
Overall	0.71 (±0.072)	0.95 (±0.022)
MC 1	0.67 (±0.113)	0.87 (±0.030)
MC 2	0.67 (±0.102)	0.89 (±0.028)
MC 3	0.44 (±0.324)	0.83 (±0.032)

CI – 95% confidence interval; MC – Modic change

Voxel-wise Modic Change Mapping

Figure 4.5 shows the [T_1 *z*-score, T_2 *z*-score] voxel-wise characterization of MCs in the training set. Cluster centroids of Modic 1, 2, and 3 were centered at [0.23 (±0.73), 1.20 (±1.16)], [1.04 (±1.00), 0.37 (±0.85)], and [-0.53 (±0.41), -0.52 (±0.85)], respectively, corresponding well with the qualitative classification system defined by hyper- and hypo-intensities. Labeling of upper and lower vertebral bodies using the rule-based classification system resulted in sensitivities of [0.67 (±0.113), 0.67 (±0.102), and 0.44 (±0.324)] and specificities of [0.87 (±0.030), 0.89 (±0.028), and 0.83 (±0.032)] for Modic

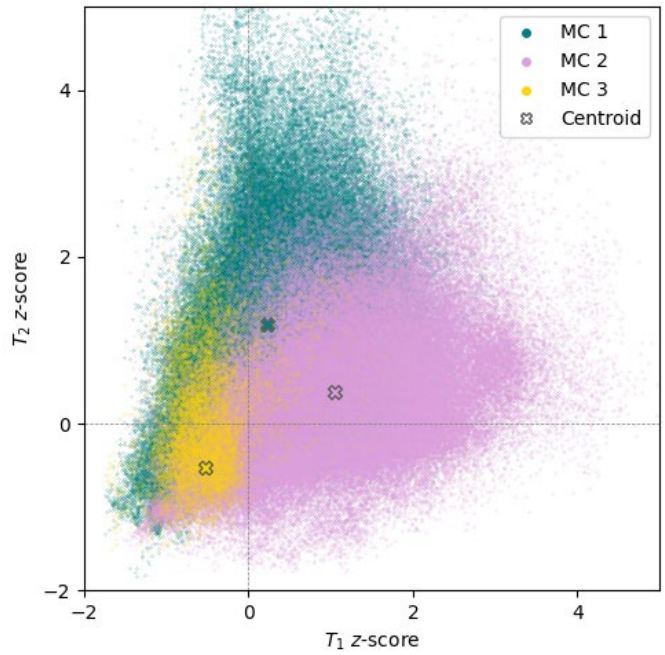

Figure 4.5 Paired T_1 and T_2 *z*-score coordinates of each voxel within Modic changes in the training set. These centroid coordinates align well with the qualitative Modic grading system and its corresponding variations in signal intensity (e.g.: Modic type 1 is hyperintense in T_2-weighted imaging, Modic type 2 is hyperintense in T_1-weighted imaging). Detected Modic changes in the test set were classified on a voxel-by-voxel basis using a nearest neighbor algorithm to these cluster centroids.

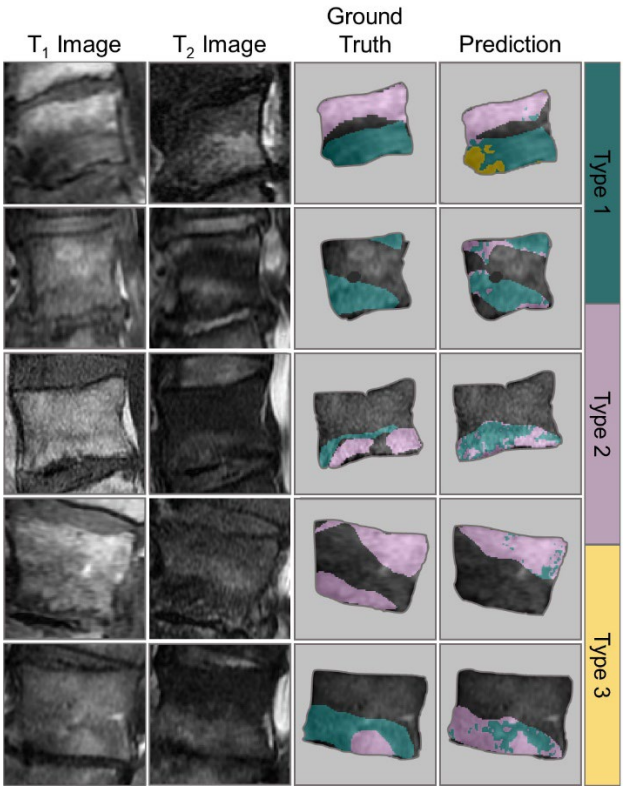

Figure 4.6 Representative examples of the model inputs (T_1 and T_2 images), radiologist-annotated ground truth segmentations, and the predicted Modic maps. The mapping technique is advantageous for visualizing heterogeneity and transitional pathology. Notably, in the top row, the model detects MC 3-like characteristics in the anterior inferior endplate. In the second row, a small MC 1 region in the anterior superior endplate, unnoticed by the radiologist, was annotated by the automatic model.

types 1, 2, and 3, respectively, as seen in **Table 4.2**. The overall prevalence of MCs in the test set was 0.27 in the ground truth and, correspondingly, 0.23 in the model predictions. Further stratification of MC prevalence is described in **Figure 4.6**. In **Figure 4.7**, representative examples of Modic maps are shown with their corresponding T_1 and T_2 images.

AI-assisted Experiment

Inter-rater agreement was initially assessed with an independently curated dataset (*n* = 20) (**Table 4.3**). Between the three radiologists, the two senior readers (reader 1 and reader 2) were in the most agreement, with a Cohen's kappa score $\kappa = 0.63$. The junior radiologist (reader 3) had moderate agreement, $\kappa = 0.52$, with reader 1 and, $\kappa = 0.45$, with reader 2.

Table 4.3 Cohen's kappa coefficients in AI-assisted experiment

	Initial Agreement (κ)	Post-AI-assist Experiment (κ)	Δκ	*p*-value
Reader 1 and 2	0.63	0.62	-0.01	NS
Reader 1 and 3	0.52	0.58	+0.06	<0.05
Reader 2 and 3	0.45	0.48	+0.03	NS

NS – not significant

With the assistance of the model prediction, agreement of reader 3 with reader 1 significantly improved to $\kappa = 0.58$ ($p < 0.05$). Agreement between readers 3 and 2 increased to $\kappa = 0.48$, though this result was insignificant by the McNemar's test. Meanwhile, reliability between readers 1 and 2 decreased slightly to $\kappa = 0.62$, again, without statistical significance.

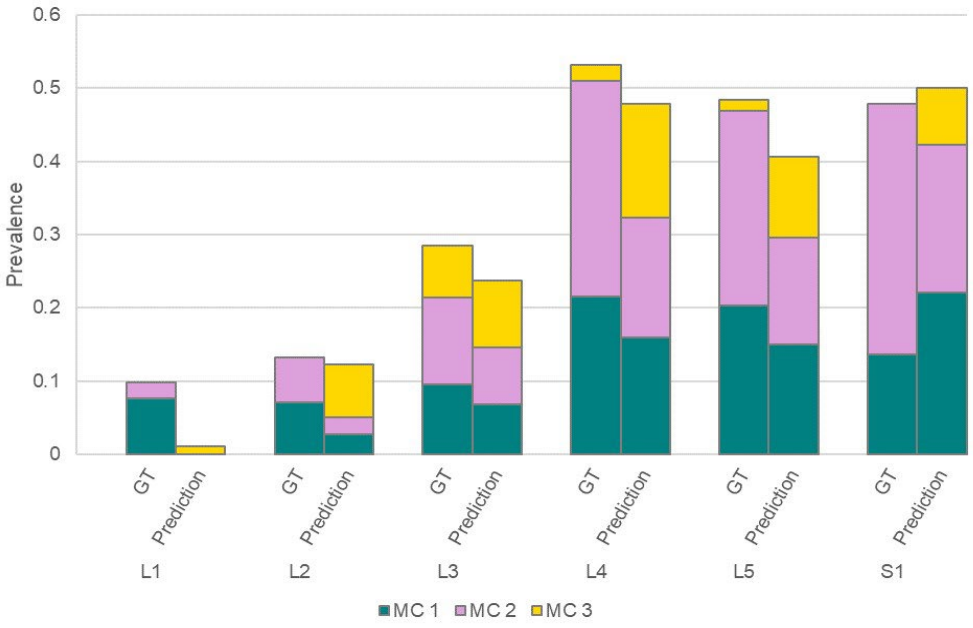

Figure 4.7 Prevalence of MCs in the ground truth and prediction of the test set, stratified by vertebral body level. The two distributions share similarities, with the highest number of MCs in the lower lumbar region (L4-S1). The prevalence is further apportioned by the relative ratios of each Modic type. The model tends to overestimate MC 3s due to low representation in the ground truth and inductive bias.

4.4 Discussion

This study used deep learning-based models to automatically localize and map MCs in vertebral bodies. Overall, these results demonstrate substantial agreement of the detection model with radiologist-annotated grading and a novel Modic mapping technique that provides grading assistance when incorporated into a radiology workflow. A design goal of this schema is to provide clinical utility through objective and interpretable models. We aimed to achieve this in two ways:

The first pertains to reducing and streamlining the semiquantitative Modic classification system into a data-driven, yet easily understood multistep algorithm. To limit the effective field-of-view to regions of the vertebral bone, rather than confounding structures such as the neighboring intervertebral discs, foramen, or spinal cord, we performed vertebral body segmentation using the V-Net [90], a widely used encoder-decoder for biomedical image segmentation. This is particularly important when considering intervertebral disc degeneration due to the strong correlation between presence of the two anomalies [155], [156]. The performance of this model is consistent with previous works in spinal segmentation [146], [148] and conveys to users of this tool which regions were evaluated by the subsequent Modic detection tool. Similarly, the rule-based classification system proposed here, based on T_1 and T_2 *z*-scores, intuitively follows the semiquantitative blueprint originally proposed by Modic, *et al.* [134]. Ultimately, the availability of intermediary results and interfaces for the pipeline's decision-making process may build confidence toward the adoption of such methodologies into clinical settings.

The second strategy adopted in this approach capitalizes on the ability of Modic maps to describe heterogeneous tissues. Systematic reviews of works involving MCs note inconsistencies in reporting procedures [136], [140]. In both research studies and in clinical practice, MCs are dictated as isolated, homogeneous lesions when they are often conglomerated and characterized by spatial heterogeneity. Past literature suggested that MRI changes may progress from Modic type 1 to type 2 to type 3 in a linear fashion [157], though recent studies have demonstrated that pathologies are often reversible [158]. Not only can MCs be transitional, it has been reported that 27.2% of MCs are regarded as mixed, comprising of characteristics of multiple Modic types [159]. Capturing the granularity of mixed MCs is challenging for the human eye, yet neural networks have proven capable of identifying detailed textural and shape features from medical imaging [160], [161]. In this work, we chose to implement a voxel-wise MC segmentation method over a classification model due to the key capability of visualizing the heterogeneity of mixed MCs. In addition, the segmentation methodology offers higher degree of supervision, where each voxel in an image is attributed with a label. This granular supervision retains context of the neighboring tissue and improves label specificity. Further works using this approach can unravel attributes of progressive or transitional MCs that may interact with LBP, as heterogeneous tissues are often correlated with degeneration.

Performance of the vertebral body segmentation and MC detection components reached or neared human reliabilities. Error analysis showed predictive inaccuracies in the lateral-most slices where partial volume effects tend to impact the delineation of bone from surrounding tissues. The performance metric is artificially deflated as the research

associate manually segmented complete vertebral bodies while the model would be apt to predict all instances of bone, some of which were only partially visible in the prescribed field of view. In the MC detection component, the distribution of predicted MCs across the lumbar vertebras were predominantly in the L4-S1 range (74.4%), which matches well with the radiologist annotations (78.8%) and past work (75.5%) [159]. Detection of MCs in L1 was notably underestimated by the model. We speculate this is due to signal loss at the periphery of the coil. Voxel-wise classification of MCs yielded high predictive value of Modic types 1 and 2, arguably the two groups most important to classify due to their prevalence [159], [162] and the strong association of MC 1 with nonspecific LBP [163], [164]. Notably, the models are trained and evaluated on a dataset with a wide arrangement of acquisition parameters to capture the variability in non-standardized imaging procedures.

In the pilot AI-assisted experiment, we found that the additional utility of the model predictions improved agreement of the junior radiologist with the senior radiologists ($\Delta\kappa = +0.06$ and $\Delta\kappa = +0.03$ with reader 1 and reader 2, respectively). However, agreement did not improve, but rather slightly decreased ($\Delta\kappa = -0.01$ with reader 1), for reassessment by reader 2. This is likely explained by the differences in training and preferences between neuroradiology and musculoskeletal radiology. The participating readers reported that a key advantage of the tool was its utility as “attention focuses,” which may have contributed to boosting agreement between reader 3 with reader 1.

The technologies developed in this study can be applied in various ways. With further development, this tool could potentially assist training efforts of junior radiologists by highlighting complex cases which depict the nuances of heterogeneous spinal

pathologies. Furthermore, because this model was trained using non-standardized clinical data, the AI-assist tool can be adapted to a continuous learning paradigm to improve model generalizability and utility without the need for additional data curation. Specifically, this model demonstrates the capability to predict transitional and heterogenous MCs which have been hypothesized to be associated with LBP. Using this tool, more data can be gathered on these changes to make consistent associations with LBP and help pave the path to elucidate the mechanisms of nonspecific LBP.

While our results demonstrate that deep learning-based approaches can contribute to identifying MCs, there are several notable limitations. First, despite the quantitative nature of this methodology, data-driven techniques are still biased by its training data and annotators. Two participants of the AI-assisted experiment were responsible for labeling the training data, which may have biased the agreement metrics against other readers. For these reasons, this algorithm is not intended to be a standalone fully diagnostic tool. Second, relatedly, we acknowledge that the exams used in this study are from a single institution and the model is not validated with multi-institutional testing. Lastly, our results are limited by the small sample size with poor representation of Modic type 3. Modic type 3 is described by signal void in both T_1- and T_2-weighted images, which makes it difficult to grade and susceptible to errors in cases with low signal-to-noise ratio. This is impactful in the nearest neighbor component of the pipeline, which is notably sensitive. Fortunately, several collaborative efforts are in-progress to amass additional data from other institutions with wider variability in imaging equipment and acquisition parameters. We also aim to extend this work by exploring domain adaptation strategies

to improve generalizability and performing longitudinal analysis to further investigate transitional pathologies.

4.5 Conclusion

In this work, we present a novel deep learning-based approach to localize and segment MCs, with results that demonstrate high agreement with radiologist grading. The introduction of this fully automatic, quantitative mapping technique may increase inter-rater reliability and ultimately improve robustness in understanding the associations of MCs with LBP and spinal degeneration.

Acknowledgements

This work is fully supported by the National Institute of Health and National Institute of Arthritis and Musculoskeletal and Skin Diseases (Project #: UH2AR076724).

Chapter 5

Multiparametric MRI Characterization of Knee Articular Cartilage and Subchondral Bone Shape in Collegiate Basketball Players

In the previous chapters, we established the utility of ML/DL models for automatically extracting imaging-derived tissue morphology. The remainder of this dissertation will focus on the harmonization of these morphologies between subjects which enables us to draw population-level conclusions.

We introduce this subject by investigating a group with high risk for injury and degeneration. Young, elite basketball players exert large amounts of force to their joints, particularly their knees. Complex repeated loading of the knee joint leads to biochemical and structural degeneration that occur prior to any visible morphological changes. In this study, we implemented a voxel-based relaxometry approach to evaluate localized compositional analyses of the knee cartilage and a statistical shape model to understand

differences in subchondral bone shape. Using these advanced computer vision methodologies, we found several characteristics unique to collegiate basketball players when compared to a low knee impact control group which may expand our understanding of local degenerative patterns in this population.

The following material in this chapter is adapted and reproduced with full permission from the publisher. The original work is referenced as:

Gao KT, Pedoia V, Young KA, Kogan F, Koff MF, Gold GE, Potter HG, Majumdar S. Multiparametric MRI characterization of knee articular cartilage and subchondral bone shape in collegiate basketball players. *J Orthop Res*. 2021 Jul;39(7):1512-1522.

5.1 Introduction

The knee is vulnerable to articular cartilage degeneration and injury in jumping athletes who exert high compressive and shear forces during practice and competitive play [1], [165]–[168]. Imparting large loads to the articular cartilage is a known risk factor for chronic musculoskeletal conditions such as early-onset osteoarthritis [165] and pain [1]. Accordingly, there is wide interest in studying associations between high-knee impact sports and long-term health of the knee joint.

Articular cartilage is of distinct concern due to its specialized function for distributing loads and its limited capacity for repair. Previous studies have used magnetic resonance imaging (MRI) to find that degenerative changes are consistently prevalent in knee cartilage of basketball players across all levels of competition [166]–[168]. A 2005 study [166] observed articular cartilage lesions on MRI in 47.5% of asymptomatic professional NBA players, with the majority of cartilage lesions found in the patellofemoral

joint. A recent study by Pappas, *et al.*, [167] imaged 24 NCAA Division I collegiate basketball players and found increased abnormal findings (fat pad edema, patellar tendinopathy, articular cartilage and meniscal injury) after one season of play in every knee imaged.

Though the high prevalence of abnormal imaging findings in high-knee impact athletes is well-established, biochemical changes of macromolecules associated with cartilage degeneration occur prior to visible morphological changes [169], [170]. Biomechanical stiffness of articular cartilage is provided by the collagen and proteoglycan (PG) organization and content, respectively, of the extracellular matrix. Damage to this macromolecular environment results in an increase of mobile water and a concomitant reduction in tissue stiffness. Compositional MRI techniques, such as $T_{1\rho}$- and T_2-relaxation time mapping, can quantify such changes in cartilage matrix biochemistry [171], [172]. $T_{1\rho}$ relaxation times reflect interactions between movement-restricted water and surrounding large macromolecules and has been related to glycosaminoglycan (GAG) and PG content and early OA. Some studies demonstrated elevated $T_{1\rho}$ relaxation times with disruption of the ECM through decreased PG content via ex vivo enzymatic removal, [173], [174] yet others have seen no relation between $T_{1\rho}$ abnormalities and GAG [175], [176]. While the mechanism is not yet fully understood, prolonged $T_{1\rho}$ has associated with populations at risk of and living with osteoarthritis [177]–[179]. Meanwhile, T_2 relaxation is associated with loss of collagen and disorganization of collagen fibrils [179]. T_2 is prolonged in the setting of degeneration of articular cartilage [174]. Newer methods permit acquisitions of $T_{1\rho}$ and T_2 in a single combined sequence and have been used to evaluate patients with anterior cruciate ligament injuries and those with osteoarthritis (OA) [179],

[180], but its use to investigate the status of knee cartilage health in young elite athletes is limited.

Quantitative analysis of $T_{1\rho}$ and T_2 relaxation time maps are traditionally performed using region of interest (ROI)-based approaches, which presents several challenges: (1) cartilage ROIs are often segmented manually or semi-automatically and are prone to inter- and intra-user variation; (2) statistical analyses are performed based on the average $T_{1\rho}$ or T_2 value of all voxels within the ROIs, limiting the spatial assessment of relaxation times within the defined regions. Methods for segmentation have recently advanced to be less reliant on manual input. Advanced segmentation methods transform images from individual knees to a single reference template, allowing comparison of local spatial distribution between subjects on a voxel-by-voxel basis. This technique, voxel-based relaxometry (VBR), has been shown to agree with ROI-based analyses [181]. Notably, it can be performed in a fully automated fashion and can provide local information and patterns of imaging markers in articular cartilage evaluation.

Another component that plays a key role in the transmission of load across the knee joint is geometric bone shape. Through skeletal homeostatic signal pathways [182], high intensity mechanical loading is associated with increased subchondral bone thickness and reduced bone resorption [182]–[184]. Stimulation of these pathways occurs in an anatomic site-specific manner depending on intensity and type of load. In turn, exercise-induced variations in bone architecture influence biomechanics of the knee joint [185], [186] and incidence rates of injury [187], [188] and osteoarthritis [188]. Due to frequent heavy loads exerted onto the knees of athletes in high-knee impact sports, it is important to classify regional bone shapes in sports with low- and high-knee impact.

Statistical shape modeling has recently gained traction as an analytical method for modeling variation in surface geometry from imaging [189], [190]. Varying algorithms have demonstrated submillimeter level matching precision, allowing for analysis of complex 3D shapes generated from medical imaging [189]–[192].

The purpose of this study was to use quantitative MRI techniques to characterize the articular cartilage and subchondral bone within the knee of two athletic groups: (1) a high-knee impact group consisting of collegiate basketball players, and (2) a non-knee impact group of collegiate swimmers. We hypothesized that the basketball players would demonstrate localized prolonged $T_{1\rho}$ and T_2 relaxation times and bone shape differences as compared to the swimmers.

5.2 Materials and Methods

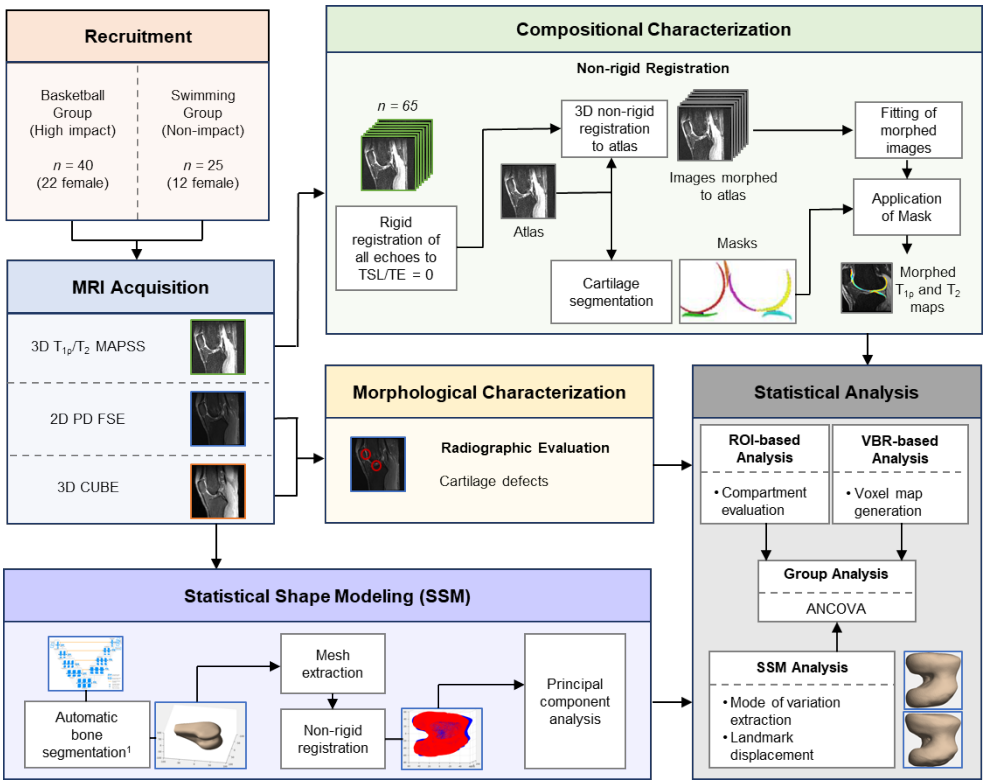

Figure 5.1 Schematic overview of multiparametric knee MRI methodology.

5.2.1 Subject Demographics

In this multicenter cross-sectional study, two cohorts of age-matched NCAA collegiate-level athletes were recruited for this study: 40 basketball players (22 female / 18 male, 19.5 ± 1.5 years, body mass index [BMI] = 24.6 ± 5.6 kg/m^2), and 25 swimmers (12 female / 13 male, 19.0 ± 1.0 years, BMI = 25.4 ± 4.9 kg/m^2). Participants were questioned about overall knee health and past history of competitive sport participation. Swimmers with prior knee injury, pain, surgery, or participation in competitive jumping sports were excluded. Procedures were performed in accordance with the rules approved by the Institutional Review Boards of the 3 participating sites. All participants provided informed written consent.

5.2.2 MRI Protocol

Imaging was performed using clinical 3T MRI (GE Healthcare, Milwaukee, WI) scanners with an 8-channel T/R (Invivo, Gainesville, FL) or an 18-channel T/R knee coil (Quality Electrodynamics, Mayfield Village, OH). Images were acquired prior to the subjects' respective basketball and swimming competitive seasons. The single-knee MRI protocol included a sagittal 2D fast spin echo (FSE) proton-density (PD)-weighted sequence, a sagittal intermediate-weighted 3D FSE CUBE sequence, and a 3D sagittal combined $T_{1\rho}/T_2$ magnetization-prepared angle-modulated portioned k-space spoiled gradient echo snapshots (MAPSS) research sequence [180]. In the $T_{1\rho}$ component of the MAPSS acquisition, time of spin-lock (TSL) was set to 0/10/40/80 *ms* using RF pulse with frequency at 500 *Hz*. Simultaneous T_2 acquisition used echo time (TE) = 0/12.8/25.7/51.4 *ms*, sharing the first image with the first $T_{1\rho}$ TSL. Additional acquisition parameters are listed in **Table 5.1**.

Table 5.1 Multiparametric MRI imaging acquisition parameters

	2D PD FSE	3D CUBE	3D MAPSS
Matrix	512 × 384	512 × 512	256 × 128
Field-of-view (*cm*)	16	16	14 or 16
Pixel bandwidth (*Hz*)	163	244	488
Slice thickness (*mm*)	3	0.7	4
Number of slices	30-45	145-210	24
Repetition time (*ms*)	5800	1200	5400
Echo time (*ms*)	40	27	0/12.8/25.7/51.4
Echo train length	14	35	1
Spin lock time (*ms*)	-	-	0/10/40/80
Flip angle	142°	90°	60°
ARC acceleration factor	-	phase: 2.0 slice: 2.0	phase: 2.0 slice: 1.0
Approximate scan time (*min:sec*)	4:30	6:30	9:40

To assess biases in quantitative measurements across the sites of acquisition, an identical phantom was imaged on all scanners. The phantom was constructed with two instances of three varying amounts of agarose to encompass a range of relaxation times and scanned with the $T_{1\rho}/T_2$ MAPSS sequence [180]. The phantom acquisition was repeated two additional times at a single site to evaluate intra-scanner variability. Coefficients of intra-scanner variation ranged from 0.2-2.2%, while coefficients of inter-scanner variation ranged from 4.1-6.6% [193].

5.2.3 Morphological Characterization

A board-certified musculoskeletal radiologist with 25 years of experience evaluated the MR images. Cartilage lesions were graded in a blinded fashion using the Modified Noyes

Score, where grade 0 classified cartilage with no lesions by PD-weighted MRI, and grades 1 and above indicated increased signal intensity or cartilage defects.

5.2.4 Voxel-based Relaxometry

Image postprocessing was performed using toolboxes implemented in MATLAB (MathWorks, Natick, MA).

For compositional analysis, all cases with cartilage lesions (modified Noyes ≥ 1) in any compartment, identified by morphological characterization, were not considered in order to focus on pre-structural abnormalities and early signs of biochemical change. Sagittal MAPSS images in all echoes were rigidly registered to the first TSL/TE = 0 of each case using VTK CISG registration toolkit [194]. Next, nonrigid registration to an atlas was then applied on all cases to morph the images to a common reference space. This was performed using elastix [195], a medical imaging registration toolbox based on maximizing mutual information between the fixed and moving images. The resulting nonrigid transformations between the atlas and each TSL/TE = 0 case was then applied to all other echoes/spin-lock images. As all images were morphed to the same coordinate space, $T_{1\rho}$ and T_2 maps were calculated on a voxel-by-voxel basis using Levenberg-Marquardt mono-exponential fitting.

5.2.5 ROI-based Relaxometry

Using a semiautomatic method based on edge-detection [196], cartilage of the atlas was segmented into 6 compartments: lateral femoral condyle (LFC), medial femoral condyle (MFC), patella (PAT), trochlea (TRO), lateral tibia (LT), and medial tibia (MT). The resulting masks were then applied to all morphed images.

A depth-dependent ROI analysis was performed to evaluate variation between cartilage layers. Each of the above-mentioned compartments were divided in half into a deep layer, closest to the subchondral bone, and a superficial layer, closest to the articular surface.

5.2.6 Statistical Shape Modeling (SSM)

Segmentation of femur, tibia, and patella bones were performed automatically using V-Net [90], a fully convolutional neural network. Bones from 36 3D CUBE images were manually segmented for training, validation, and testing (26/6/4 split). Prior to training, all images were downsampled to 256×256×212 for computational efficiency. All training images were augmented with a random permutation of the following preprocessing techniques: additive gaussian noise, histogram matching, gaussian filter, and affine transformation.

The V-Net architecture implemented 8 output channels in the first level, doubling at each of the subsequent 3 levels. One, 2, and 3 convolutions were performed at each level, respectively, and 3 additional convolutional layers were added to the bottom level of the network. Dropout was implemented at 5% as a regularization penalty. Dice coefficient was chosen as the loss function, with sigmoid activation applied, as well as the metric for evaluation. The model was trained for 24,000 iterations using a batch size of 1 and resulted in Dice coefficients of 0.98 ± 0.01, 0.98 ± 0.01, and 0.96 ± 0.01 (mean $\pm$ SD) for the femur, tibia, and patella, respectively. The prediction algorithm was then applied to each case in the dataset.

The resulting segmentations were used to produce 3D triangulated meshes of the femur, tibia, and patella bones using a Marching Cube algorithm [197]. Next, with the

bones in all cases being represented by clouds of points, each bone was non-rigidly registered using FOCUSR, as proposed by Lombaert, *et al*. [198], This method utilized spectral correspondence, which parametrizes vertex similarity using Laplacian eigen-decomposition and then performs spectrum reordering via feature matching. The registered femurs, tibias, and patellas were described with 50537, 33210, and 8477 vertices, respectively.

Principal component analysis (PCA) was then performed to simplify the complexity of the surface data for interpretation. PCA transformed the vertex coordinates to orthonormal bases, where each principal component (PC) mode is uncorrelated and is ordered such that the first PC describes the direction of maximal bone shape variance and subsequent PCs are sorted in a descending manner. In consideration of the size of our dataset, 10 PC modes were sufficient to capture over 80% of variance in each bone while still maintaining physical interpretability of the surface models.

5.2.7 Statistical Analysis

Morphological statistical analysis used a chi-square test to assess the relationship of the prevalence of cartilage abnormalities between the two groups.

In compositional analysis, summary statistics, including mean and standard deviation (SD), of $T_{1\rho}$ and T_2 times were computed and compared between basketball players and swimmers. This was computed for individual voxels in VBR, and in each cartilage compartment and between cartilage layers in ROI-based analysis. Group differences were assessed using one-way analysis of covariance (ANCOVA). Gender, BMI, and site of acquisition were used as adjustment factors to control for confounding

effects. A significance threshold was set at $p < 0.05$ (SPSS version 26.0, IBM, Armonk, NY).

Bone shape analysis involved evaluating PC values to determine if specific shapes were associated with the basketball or swim group. PCs that described shape differences related to the femur and tibial shafts were disregarded, due to variations in subject positioning during MRI acquisition and our specific interest in characterizing subchondral bone. An ANCOVA test, controlled for gender, BMI, and site of acquisition, determined statistical difference between groups. Physical representation of each mode was visualized in two ways: (1) average surface ± the displacement of each vertex by 3 standard deviations (SD), and (2) the average surface with color mapping of the Euclidean norm at ±3 SDs.

5.3 Results

5.3.1 Morphological Evaluation

The prevalence of cartilage abnormalities was significantly higher in the basketball group ($\chi^2 = 6.658$, $p < 0.01$), occurring in 24.6% of knees of basketball players and 6.3% of knees of swimmers (**Table 5.2**). By compartment, this increase was significant in the LFC ($\chi^2 = 5.51$, $p < 0.05$).

5.3.2 ROI Analysis

Sixteen of 65 basketball cases and 3 of 48 swim cases included one or more defects in any cartilage compartment and were removed from ROI- and subsequent VBR-based analysis to isolate differences in tissue composition. Mean $T_{1\rho}$ and T_2 values of the compartmentalized results ranged from 34.3-46.3 *ms* and 25.0-32.9 *ms*, respectively.

Table 5.2 Counts of cartilage abnormalities

		Basketball (65 knees)		Swim (48 knees)		χ	*p*-value
LFC	Noyes 0	58	(89.2%)	48	(100%)	5.51	Significant
	1	1	(1.5%)	0	(0%)		$p < 0.05$
	2	6	(9.2%)	0	(0%)		
MFC	Noyes 0	64	(98.5%)	47	(97.9%)	0.47	Insignificant
	1	0	(0%)	1	(2.1%)		$p = 0.83$
	2	1	(1.5%)	0	(0%)		
TRO	Noyes 0	63	(96.9%)	48	(100%)	1.50	Insignificant
	1	1	(1.5%)	0	(0%)		$p = 0.22$
	2	1	(1.5%)	0	(0%)		
PAT	Noyes 0	56	(86.2%)	46	(95.8%)	2.94	Insignificant
	1	2	(3.1%)	0	(0%)		$p = 0.09$
	2	6	(9.2%)	2	(4.2%)		
LT	Noyes 0	63	(96.9%)	48	(100%)	1.50	Insignificant
	1	0	(0%)	0	(0%)		$p = 0.22$
	2	2	(3.1%)	0	(0%)		
MT	Noyes 0	65	(100%)	48	(100%)	--	--
	1	0	(0%)	0	(0%)		
	2	0	(0%)	0	(0%)		
Total Knees (Noyes ≥ 1)		16	(24.6%)	3	(6.3%)	6.66	Significant $p < 0.01$

The ROI-based results demonstrate significant group differences in the medial compartment (**Figure 5.2**). The basketball group had significantly prolonged $T_{1\rho}$ and T_2 values in the medial femoral condyle ($T_{1\rho}$: 3.54% difference, $p < 0.001$, and T_2: 3.63% difference, $p < 0.001$) and medial tibial ($T_{1\rho}$: 5.28% difference, $p < 0.001$, and T_2: 6.04% difference, $p < 0.001$) compartments, as well as prolonged T_2 values in the lateral femoral

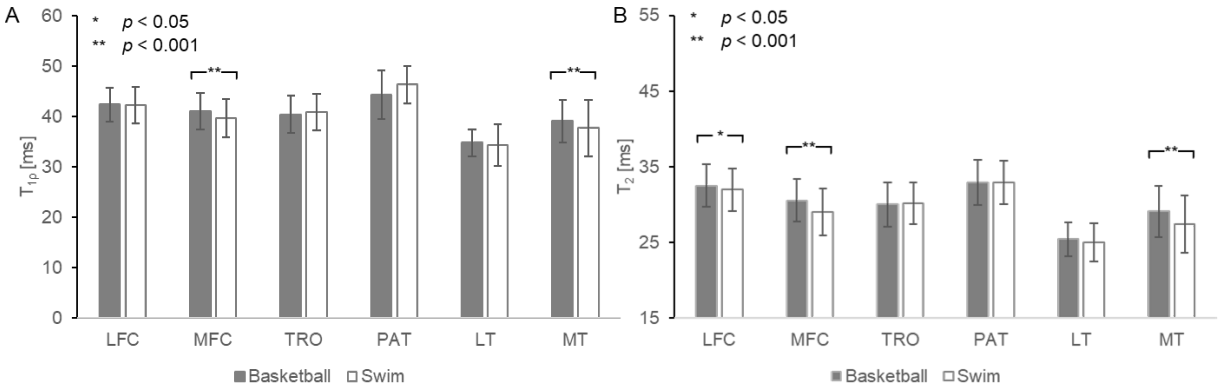

Figure 5.2 Results of the ROI-based method for analysis of group differences in (A) $T_{1\rho}$ and (B) T_2 within cartilage compartments. Significant differences were found in the MFC and MT in both $T_{1\rho}$ and T_2, and additionally in the LFC in T_2.

condyle compartment (T_2: 1.72% difference, $p < 0.05$), though this lateral association was weaker. No significant differences were detected in the patellofemoral compartment ($T_{1\rho}$: 4.30% difference, $p = 0.53$, and T_2: 0.17% difference, $p = 0.43$) using the ROI-based technique.

Cartilage compartments were further partitioned into a deep and superficial layer to evaluate variations in cartilage depth. When comparing these laminar features in each compartment, $T_{1\rho}$ and T_2 of the superficial layer was significantly prolonged ($p < 0.001$ in all cases). Group analysis, shown in , demonstrated similar results to those in **Figure 5.2** prior to partitioning, with statistically prolonged $T_{1\rho}$ in the deep layer of the LFC (2.36% difference, $p < 0.001$), both layers of the MFC (deep: 6.05% difference, $p < 0.001$; superficial: 6.84%, $p < 0.001$) and MT (4.78% difference, $p < 0.05$; superficial: 5.21% difference, $p < 0.001$) in basketball players. T_2 was similarly prolonged in basketball players for most cartilage compartments: LFC (deep: 4.29% difference, $p < 0.001$; superficial: 1.76% difference, $p < 0.05$), MFC (deep: 6.64% difference, $p < 0.001$;

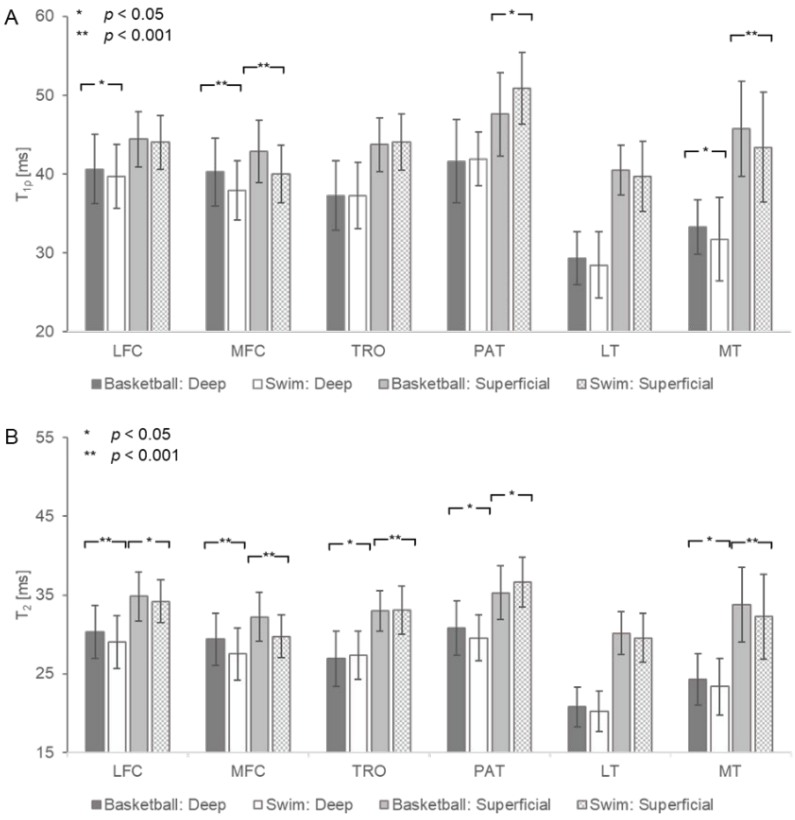

Figure 5.3 Depth-dependent group comparison of mean $T_{1\rho}$ and T_2 relaxation times. In sub-compartments with statistical significance, basketball players demonstrate prolonged relaxation times except in the superficial layer of the patellar and trochlear cartilage.

superficial: 7.94% difference, $p < 0.001$), PAT (deep: 4.09% difference, $p < 0.05$) and MT (deep: 3.97% difference, $p < 0.05$; superficial: 4.56% difference, $p < 0.001$). The only sub-compartment where relaxation times of swimmers were higher than those of basketball players was the superficial layer of the patellar ($T_{1\rho}$: 6.68% difference, $p < 0.05$; T_2: 3.77% difference, $p < 0.05$), and trochlear cartilage (T_2 deep: 1.52% difference, $p < 0.05$; T_2 superficial: 0.20% difference, $p < 0.001$).

5.3.3 VBR Analysis

Interpretation of the mean $T_{1\rho}$ and T_2 maps from VBR displayed prolongation near the trochlear groove and areas of shortening in the anterior and posterior regions of the tibio-femoral articulation.

Comparison of the two groups demonstrated significant differences by sport, with basketball players generally with longer $T_{1\rho}$ and T_2 values, particularly in both femoral condyles (lateral: 12.63% average % difference, 42.7% significant voxels; medial: 3.48% average % difference, 29.2% significant voxels) (**Figure 5.4**). The voxels that depicted significant prolongation were heavily focused in the posterolateral and posteromedial femur. Diffuse elevation was also noted in the antero-cartilage.

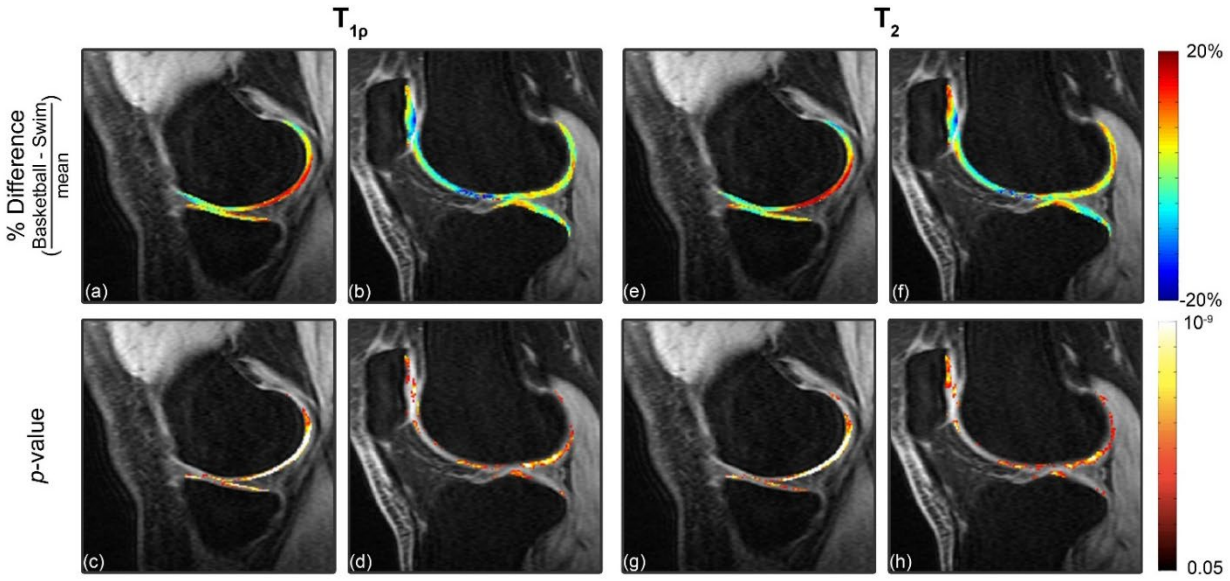

Figure 5.4 Representative examples of VBR group analysis. (a)-(b) Percentage difference map ($\frac{\bar{T}_{,basketball}-\bar{T}_{,swim}}{\frac{\bar{T}_{basketball}+\bar{T}_{swim}}{2}}$), where $\bar{T}$ is mean relaxation time, and (c)-(d) the respective p-maps generated by VBR. Basketball players have significantly prolonged $T_{1\rho}$ values in the posterior medial and lateral femoral condyles and tibial plateau. Meanwhile, there are laminar differences in the deep and superficial layers of the patellofemoral joint. Basketball players present with prolonged $T_{1\rho}$ values in the deep layer and shorter $T_{1\rho}$ values in the superficial layer, as compared to swimmers. (e)-(h) T_2 difference and *p*-maps show similar differences between groups.

Voxel-based group analysis also revealed differences through the depth of the articular cartilage in the patellofemoral joint: basketball players had higher $T_{1\rho}$ and T_2 values in the deep layer of cartilage while swimmers had prolonged values in the superficial layer. This depth-dependent distribution was not obviously evident in other regions evaluated and is of notable interest due to the vital role of the patellofemoral joint in the translation of weight.

5.3.4 Bone Shape Analysis

The femur, tibia, and patella were each described in domains defined by 10 PC modes which maximize variation in shape. The amount of variability within the entire dataset, as represented by the PCs were 80.8%, 89.7%, and 82.5%, respectively.

Among the 10 PCs of each bone, ANCOVA tests showed 3 total modes that were significantly different between groups (**Figure 5.5**): the 2nd and 7th modes of the tibia, and the 4th mode of the patella. Tibia mode 2 ($p < 0.01$, 22.0% of variance) describes size of the lateral plateau relative to the medial plateau, particularly in the anterolateral aspect. Tibia mode 7 ($p < 0.05$, 1.87% of variance) represents the relative heights of the intercondylar eminence. Patella mode 4 ($p < 0.01$, 4.24% of variance) is related to the curvature and convexity of the lateral articular facet. Variance in this mode also seemed be connected to patellar symmetry. As the lateral facet extended, in relation to the medial facet, it demonstrated increased convexity.

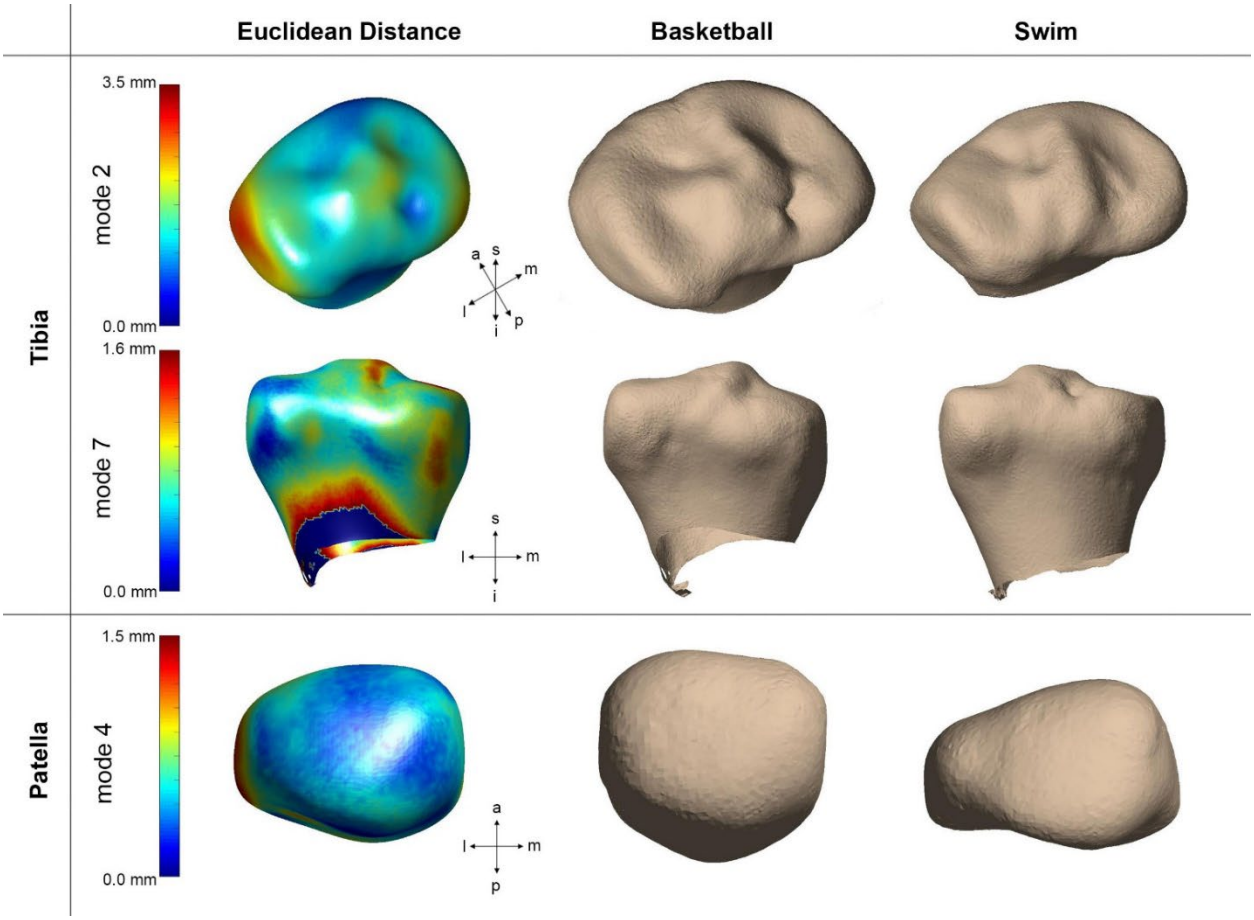

Figure 5.5 Bone shapes demonstrated significant group differences in the tibia and patella. In the first column, an average surface mesh is mapped in color by the Euclidean distance between the average surface mesh and +3 SD. Models of basketball players and swimmers are shown in the second and third columns, respectively, represented by mean vertices displaced by ±3 SD. (Top) Tibia mode 2 relates the size of lateral plateau, (middle) tibia mode 7 primarily describes to relative heights of intercondylar spines, and (bottom) patella mode 4 show symmetry and curvature of medial and lateral facets.

5.4 Discussion

This multicenter study used multiparametric MRI to extensively characterize the articular cartilage and bone shape of knees of basketball players (high knee-impact) and swimmers (no knee-impact). We demonstrated significant group differences using morphological evaluation, compositional evaluation through a traditional ROI-based and

a fully automatic VBR-based techniques, and statistical subchondral bone shape comparison in the femur, tibia, and patella.

Imaging abnormalities in the articular cartilage of basketball players has been a heavily studied area [1], [165]–[168]. Frequent jumping, running, pivoting/cutting motions applies heavy mechanical load to the cartilage. The prevalence of morphological cartilage defects found in this study relates well to past findings in imaging studies of professional basketball players [166], [168]. Distribution of defects across the cartilage compartments was also consistent, with remarkably high pervasiveness of findings in the patellar cartilage, followed by the trochlear and femoral cartilage.

MRI studies of knee cartilage composition in basketball players is much more limited. In this study, we identified and removed knees with morphological abnormalities from analysis to highlight key differences in biochemistry between groups. Classic ROI segmentation and analysis of $T_{1\rho}$ and T_2 led to findings of significant group differences in the medial compartments. Recent *in vivo* experiments of compartmental strain of the tibia show increased strain on the medial side of the tibial plateau with increased normalized walking speed, but not for the lateral side [199]. Additionally, medial compartment OA is the most common form of OA [200], [201]. Our ROI-based analysis captured a pattern representative of this asymmetry; however, this method was not effective in finding local findings in other compartments. The traditional ROI-based analysis detected no differences between groups for patellar cartilage, despite previous research indicating prevalence of imaging findings in this compartment [166], [168].

Overall, the VBR analysis was more sensitive to local differences. The prevalence of significant $T_{1\rho}$ and T_2 prolongation in the medial femur and medial tibia was consistent

with results found in ROI analysis and was characterized by a dominance of higher values in the posterior cartilage of the basketball group. Depth dependent differences were detected, as well. Basketball players demonstrated higher values in the superficial layer of the medial cartilages, as seen in both VBR and depth-dependent ROI results. However, an opposite pattern was displayed on the lateral side where diffuse patches of higher relaxation times are evident in the deep layer. The superficial layer of cartilage is composed primarily of type II and IX collagen, aligned parallel to the surface to protect the deep layer from shear stress, while the deep layer contains higher proteoglycan content and collagen aligned perpendicular to the surface to resist compressive forces [202]. The VBR results may suggest cartilage degeneration in areas of prolonged $T_{1\rho}$ and T_2. We speculate, these differences could be attributed to the complexity of joint loading, and the differences in mechanical loading between the two sports: the basketball group experiences relatively more compressive loading on the lateral side, possibly from pivoting/cutting motions and high magnitude jump-landing. Similarly, the dichotomy of relaxation patterns in the patellar cartilage could possibly be due to its role in facilitating extension during jumping and squatting. Conversely, swimmers use high frequency, low magnitude flexion/extension movements, which may exert higher shearing and tensile forces on the superficial layer of the patella and trochlea. Clearly, further experiments are warranted in support of these mechanistic hypothesis.

External loading is known to influence subchondral bone shape and thickness via bone remodeling [184], [185]. While our SSM results demonstrated no differences in femur shape, the significant modes of the tibia are especially relevant in controlling the biomechanics of the tibiofemoral joint. We found more symmetry between lateral and

medial plateaus in basketball players as compared to swimmers. Functionally, the lateral plateau is convex in shape and performs translational motion to the concave medial plateau. The anterolateral plateau, specifically, experiences tibial subluxation during knee flexion, indicating tibio-femoral internal rotation [203]. High degree of rotation due to pivoting/cutting in basketball may contribute to the symmetry seen in the lateral plateau shape. Similarly, tibia mode 7 shows higher prominence of the medial spine in basketball players. With its physical connection to the anterior cruciate ligament and its proximity to the medial meniscus, a vital tissue in shock absorption, the asymmetric heights could be explained by increased mechanical loading and subsequent bone remodeling. The size of the tibial plateau [204] and heights of the intercondylar eminence have been positively correlated with the prevalence of tibial osteophytes [205] and osteoarthritis [206]. Therefore, tibia shape may be an important consideration in identifying progression of knee kinematics, degeneration, and risk of injury in young athletes.

In regard to patella shape, there was significant variation in the lateral facet between groups. The representative patella of basketball players was more symmetric with a concave lateral facet, whereas that of swimmers was elongated and convex. Using Wiberg shape classification [207], the patellar shape of basketball players can be categorized as type I, showing congruency and concavity on both facets of the patella. Meanwhile, the non-impact group shows similarity to type III, with a convex and posterior-sloping medial facet much smaller than the lateral facet. It is unclear how this shape is associated with swimmers, as there may be hidden confounding factors that were not accounted for in this study.

In summary, we identified several characteristics associated with high-knee impact athletes, including prolonged $T_{1\rho}$ and T_2 relaxation in cartilage compartments and local depth-dependent differences, as well as bone shape variations in the tibia and patella. This study had several limitations. First, morphological evaluation was performed by a single senior musculoskeletal radiologist. Second, the voxels generated with the 3D MAPSS sequence were large (0.6 mm × 1.2 mm × 4 mm) as compared to the cartilage thickness, a known intrinsic MRI limitation when balancing the factors of scan time, anatomic coverage, and voxel size. Even with the possible influence of partial volume artifacts, our results show significant statistical results sensitive to local distributions. This is particularly evident in the patella, where cartilage is thicker. Finally, the playing season training regimen of the basketball players and swimmers was not incorporated into the current analysis. Further evaluation of this dataset may incorporate the effects of position played, for basketball, or primary stroke, for swim. It is worthwhile to note that the findings of this study do not establish causation between play and $T_{1\rho}$ or T_2 prolongation or bone shape. This cross-sectional study evaluates absolute quantitative measures at a single timepoint; future evaluations will incorporate the effect of one season of play and longitudinal changes in these populations to further establish our findings. The relationships described in the current study provide a comprehensive characterization of the knees of young athletes with considerably different loading patterns, derived from imaging alone.

ACKNOWLEDGEMENTS

This work was supported by the NBA and GE Healthcare Orthopedics and Sports Medicine Collaboration.

Chapter 6

Large-Scale Analysis of Meniscus Morphology as Risk Factor for Knee Osteoarthritis

One of the major draws of DL-based pipelines is the ability to automatically process unprecedentedly large amounts of data with improved generalizability. We have seen a commensurate curation of big datasets in medical imaging. In the osteoarthritis field, the Osteoarthritis Initiative (OAI) [208] has been foundational in large-scale analyses, following 4,796 participants across 8 years of clinical and imaging visits. Researchers in our group have developed and validated several tissue segmentation models for MRIs in the OAI dataset. Please see their work: [209], [210], and [211]. These efforts have enabled surveys of morphological OA features on a scale unlike we have seen before.

In this chapter, we describe how to integrate DL segmentation and SSM to draw population-level conclusions about tissue geometry. More specifically, we utilize

computer vision tools to investigate the geometric shape of the meniscus in the knee joint and determine their relationship to future OA onset. However, as alluded to in Section 2.1.2, knee OA is a whole joint disease. Thus, we leverage the cartilage and bone segmentations to spatially encode rates of change in cartilage thickness to find relationships between meniscus shapes and cartilage degeneration.

The following material in this chapter is currently in press at *Arthritis & Rheumatology*. The reference is below:

Gao KT, Xie E, Chen V, Iriondo C, Calivá F, Souza RB, Majumdar S, Pedoia V. Large-scale Analysis of Meniscus Morphology as Risk Factor for Knee Osteoarthritis. Accepted by *Arthritis Rheum*. 2023.

6.1 Introduction

One of the leading causes of global disability [212], knee osteoarthritis (OA) is a whole joint disease with complex, multifactorial pathophysiology [45]. The growing recognition of interactions between structural tissues of the knee joint have pushed the research community to categorize distinct OA phenotypes using biochemical and imaging findings [213]. Traditionally, articular cartilage degeneration and changes in subchondral bone have been established as the foremost biomarkers of disease; however, the meniscus has garnered attention for its integral role in dispersion of load to the articular cartilage and overall mechanical stability of the joint [214], [215]. Damage to the meniscus, particularly tears and tibial extrusion, are strongly associated with knee pain, cartilage loss, and OA progression [216]–[221]. Several widely disseminated semiquantitative

imaging scoring systems have acknowledged meniscal injury as one of the main structural phenotypes in the standardized assessment of knee OA [61], [62], [222].

While links between meniscus injury and knee OA are well established, the characterization of the geometric shape of the meniscus and its relationship with knee OA remains a challenging area of research. Past efforts identify and manually measure meniscal shape features hypothesized to be associated with OA risk factors. Wirth, *et al.* [223], segmented the meniscus from 31 proton density-weighted magnetic resonance images (MRIs) of knees with and without radiographic OA and found that knees with OA had greater meniscus volume, surface area, thickness, and increased extrusion. Kawahara, *et al*. [224], similarly extracted measurements from menisci of 51 subjects using manual segmentation of $T_{1\rho}$-weighted MRI. Knees in the severe OA group were characterized with larger longitudinal diameter and posterior wedge angle and smaller posterior wedge width of the medial meniscus. Recently, Wenger, *et al.* [225], discovered that both the medial and lateral menisci of OA knees bulged at the periphery and were more extruded, and additionally, the lateral menisci were larger in volume. Exploration of meniscus geometry has uncovered aspects of the relationship between coverage, disruption of cartilage, and OA but may be insensitive to nuanced shape features beyond the engineered constructs specified in the abovementioned studies. Moreover, due to the complexity of meniscus morphologies, there is a need to evaluate shape on a larger scale.

The recent surge of deep learning-based segmentation techniques has impacted various areas of imaging-derived morphological analyses, including those of knee OA. The U-Net [85] and other convolutional neural network models have been shown to delineate cartilage, bone, meniscus, and other soft tissues with high reproducibility,

equivalent to or surpassing human inter-operator variability [226]–[232]. Specialized techniques have pushed the performance of knee joint segmentation models, including anatomical shape-assisted model training [226], self-attention mechanisms [227], and adversarial learning schemes [229]. Irrespective of the underlying components, deep learning-based models have quickly become pervasive for their ability to automatically output high-quality segmentation of complex structures.

The capacity to expeditiously extract and analyze tissue morphology with deep learning has enabled advancements in large-scale shape analysis. Benefitting greatly from increased sample sizes, statistical shape modeling (SSM) is a technique that parameterizes and compactly describes population-level geometric features [233]. The methodology, grounded in computer vision, commonly begins by algorithmically matching anatomical landmarks between subjects. By forming this inter-subject correspondence, the geometric variations within a population can be quantified and then condensed into a hierarchy of major modes of deviation from the average shape. A key advantage to SSM is the aspect of statistical parameterization, which allows for the identification and reconstruction of tissue geometries representative of subpopulations. Several shape models have been developed for the knee joint to describe variations in bone shape such as intercondylar narrowing in subjects with acute anterior cruciate ligament injuries [189], prominence of the medial tibial spine in athletes with high-knee impact [234], and classical structural signs of disease progression in OA patients [190], [235], [236]. Notably, Bowes, *et al.* [236], established a quantitative SSM measure of bone shape to reflect OA status, a step towards the development of personalized shape metrics. While limited by small sample sizes, pilot shape models have been developed for the meniscus as well [237],

[238], typically utilizing manual or semiautomatic techniques. Automated shape extraction and data-driven analysis powered by expanded sample sizes capable of capturing natural shape variability would greatly benefit the discovery of the relationships between meniscal morphology and OA onset or progression.

In this study, we leveraged deep learning-based segmentation of multiple tissues in the knee joint to perform shape analysis of the meniscus. We constructed an SSM of the meniscus with the following aims: (1) to identify meniscus shapes associated with future onset of OA, and (2) to localize future changes in cartilage thickness with respect to at-risk meniscus shapes. This automated methodology was scaled for evaluation of the Osteoarthritis Initiative (OAI), a multicenter, prospective, observational imaging dataset sponsored by the National Institutes of Health.

6.2 Materials and Methods

This analysis utilized baseline MRI acquisitions from the OAI dataset to generate a statistical model of meniscus shape. Shape features associated with future development of OA were identified via group analysis between a healthy *Control* group and an *OA Incidence* group. Each baseline meniscus shape was subsequently evaluated with longitudinal cartilage thickness changes. The study design overview is outlined in **Figure 6.1**.

6.2.1 Dataset

Imaging data was obtained from the 4,796 participants with, or at risk for, symptomatic femoral-tibial knee OA enrolled in the OAI study. Knee radiographs and MRIs (Siemens

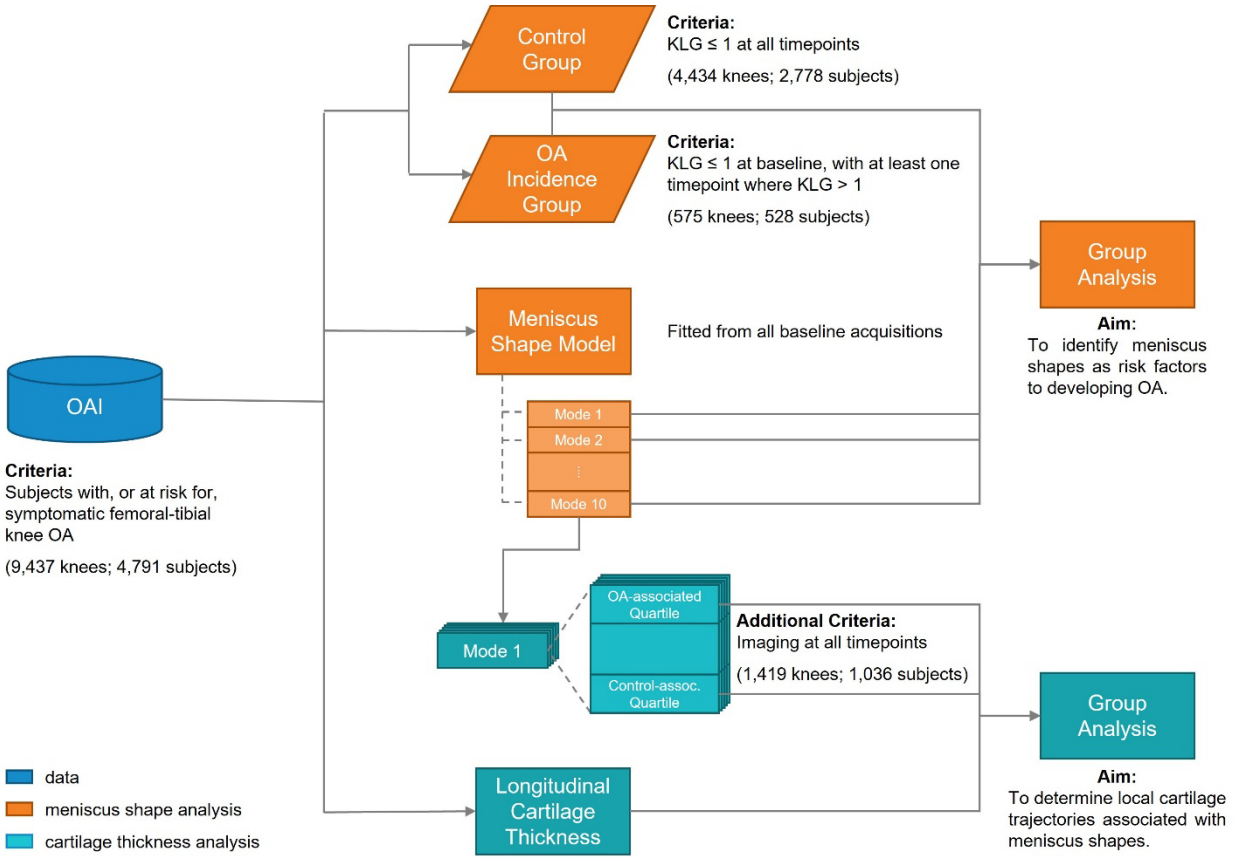

Figure 6.1 Meniscus shape study design. This analysis of meniscus shape utilized 4,790 DESS-we MRI acquisitions from the OAI dataset. A statistical shape model was generated for the meniscus and shape scores between the *Control* and *OA Incidence* groups identified features associated with future OA onset. These features were then assessed with localized cartilage thickness changes to investigate the link between meniscus shapes and future cartilage degeneration.

Trio 3.0 Tesla) were acquired annually at baseline to 48-month visits, and every two years between 48-month and 96-month visits [239], [240].

Kellgren-Lawrence (KL) grades of OA severity were performed centrally using bilateral PA fixed flexion knee radiographs to describe cases with osteoarthritic signs of cartilage and bone such as joint space narrowing and osteophytes. Meanwhile, MRIs were graded with the MRI Osteoarthritis Knee Score (MOAKS) system by the Boston Imaging Core Lab, encapsulating morphological and signal abnormalities of the cartilage, bone, meniscus, ligaments, and tendons.

Table 6.1 Demographic information of OAI participants and study subsets

Dataset	Number of Participants	Number of Knees	Gender (*M/F*)	Age (*years*)	Height (*m*)	Weight (*kg*)	BMI (kg/m^2)
Total	4,790	9,418	1,991 / 2,799	61.2 (±9.2)	1.68 (±0.097)	81.3 (±16.4)	28.6 (±4.8)
Normal at Baseline	3,105	5,017	1,360 / 1,745	60.3 (±9.2)	1.69 (±0.096)	79.2 (±15.9)	27.8 (±4.5)
Control	2,778	4,434	1,242 / 1,536	60.2 (±9.2)	1.69 (±0.096)	78.7 (±15.9)	27.6 (±4.5)
OA Incidence	528	575	178 / 350	60.4 (±8.7)	1.67 (±0.097)	81.7 (±15.7)	29.1 (±4.6)
Cartilage Thickness Subset	1,036	1,419	455 / 581	59.1 (±8.6)	1.69 (±0.094)	78.5 (±15.7)	27.4 (±4.3)

*Continuous variables are described by the mean and standard deviation.

9,418 knee MRIs at the baseline timepoint from 4,790 participants were used to build the meniscus shape model. Six participants were excluded due to unavailable imaging. Statistical analysis considered 5,009 knees from 3,103 participants with no OA (KL grade ≤ 1) at baseline. In the longitudinal cartilage thickness analysis, 1,419 knees from the 1,036 participants with complete MRIs at all seven timepoints were included. Demographic data of all participants and the abovementioned subsets can be found in **Table 6.1**.

6.2.2 Image Processing

The summary of the technical methodology is illustrated in **Figure 6.2**.

Tissue Segmentation

With high spatial resolution and excellent delineation of the cartilage-bone and cartilage-meniscus interfaces [240], sagittal 3D dual-echo steady-state images with selective water excitation (DESS-we) were automatically segmented to extract tissue morphology. This

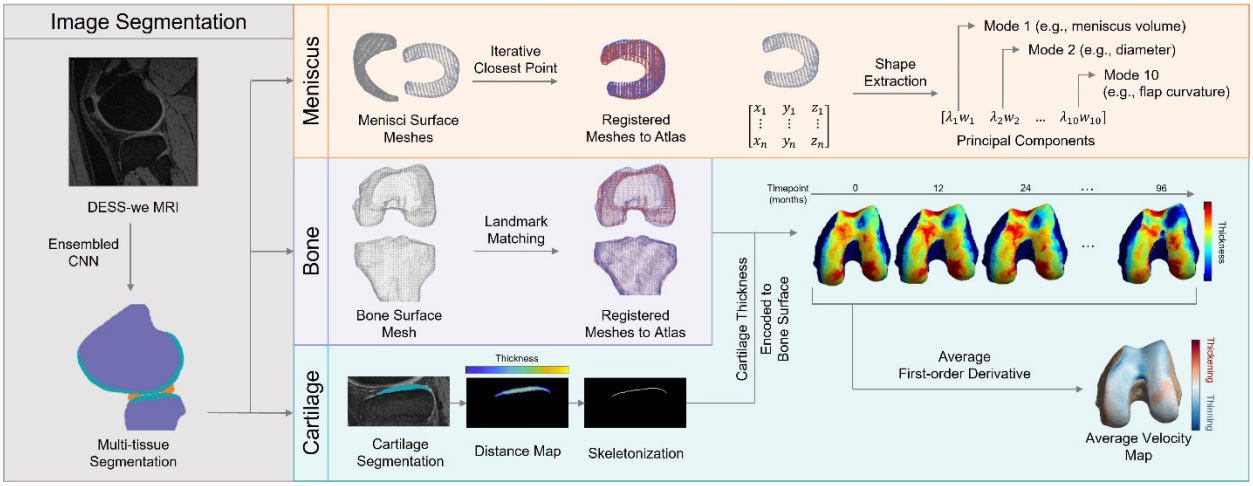

Figure 6.2 Schema of technical methodology. Menisci, bone, and cartilage are segmented from DESS-we MRI. The menisci are transformed into surface meshes and registered to the menisci of a selected atlas, aligning all cases to a common coordinate system. Using PCA, geometric shape features of the meniscus can be extracted from the registered surface meshes. Meanwhile, segmentations of the femur and tibia are similarly characterized. Points on the subchondral bone are further encoded by cartilage thickness and average velocity values.

was performed using neural network-based models that have been trained and validated for the OAI in previous efforts [209], [210]. Performance of these models were measured using mean (standard deviation [SD]) Dice coefficients to describe average overlap between human-annotated and model-predicted segmentations: meniscus = 0.874 (±0.024), femur = 0.972 (±0.011), tibia = 0.973 (±0.013), femoral cartilage = 0.890 (±0.023), and tibial cartilage = 0.880 (±0.036).

Automatic Landmark Correspondence

Next, the segmentation masks of the meniscus and bones were transformed to triangular meshes using the Marching Cubes algorithm [197]. This effectively reduced the complexity of tissue morphology to its surface topology.

A healthy knee within the dataset was chosen to serve as an atlas. The criteria for atlas selection were as follows: (1) KL grade = 0, (2) no meniscal abnormalities by

MOAKS, (3) no imaging abnormalities, and (4) minimal deviation from mean age and BMI of the dataset. After flipping knees of opposite laterality (i.e., right versus left) to match the atlas, all surface meshes were then algorithmically aligned. For the menisci, this included scaling to the size of the atlas menisci, rigid alignment using Iterative Closest Point, and re-scaling to the original size. Due to the prominence of anatomical bony features, the femur and tibia were instead independently aligned using a landmark matching algorithm [198]. With the atlas-matched correspondence, geometric features of the surface topology could be compared in the downstream statistical analysis.

Statistical Shape Modeling of the Meniscus

Principal component analysis (PCA) of the matched meniscal meshes served to reduce the dimensionality of the feature-spaces from n_m and n_l to k, where n_m and n_l are the numbers of x, y, and z coordinates of the vertices in the medial and lateral menisci surface meshes, and k is a designated number of representative shape features, known as modes of variation [189]. To highlight shapes as risk factors for OA, only meshes of baseline visits were considered to build the initial n_m- and n_l-dimensional spaces. Each mode was then assessed by simulating variation around the mean shape to assign a morphological description to the quantitative variance.

Localization of Cartilage Thickness Trajectories

Cartilage thickness values were encoded to the subchondral surfaces of the bone mesh. Thicknesses were derived using a Euclidean distance transform of the binary cartilage segmentation mask on a slice-wise basis. The distance maps were skeletonized along the long axis of each cartilage and encoded to the nearest vertex of the surface mesh.

Longitudinal changes in cartilage thickness, *i.e.*, thickness velocities, were computed for participants with complete MRI across the seven timepoints (baseline to 8 years). Velocities at each vertex of the cartilage skeletons were independently computed as the first-order derivatives of the observed thicknesses and averaged to estimate future localized thinning or thickening from baseline.

6.2.3 Statistical Analysis

To determine meniscus shape features associated with future OA incidence, two groups were established: (a) *Control*: knees with KL grade of 0 or 1 throughout participation in the OAI, and (b) *OA Incidence*: knees with KL grade of 0 or 1 at baseline and subsequent incidence of radiographic OA (KL grade $\geq$ 2) within 8 years. Group analysis was performed using one-way analysis of covariance and least-squares mean, controlling for age, sex, race, and body mass index (BMI). Statistical significance was established as $p < \frac{0.05}{k}$ to correct for independent comparisons of each mode.

It is plausible that meniscal damage, such as tears and extrusion, are linked with meniscal shapes identified as OA risk factors. To investigate this assumption, knees in the *Control* and *OA Incidence* groups with complete MOAKS grading for meniscal anterior, body, and posterior tears, posterior root tear, and anterior extrusion were fitted in a multiple linear regression using at-risk meniscal shapes, the aforementioned MOAKS grades, and demographics (age, sex, race, and BMI) as predictors for OA incidence.

Additional analysis was performed to establish cross-sectional patterns in meniscus shapes with relevant demographic subpopulations. One-way analysis of covariance was used to evaluate differences in gender, while Pearson correlation

assessed associations with age, height, weight, and BMI. Demographic values were extracted from the baseline visit.

Furthermore, the prognostic capacity of meniscus shapes and their relationship to future development of common imaging abnormalities were assessed. These included tears of the anterior horn, meniscal body, posterior horn, posterior root, and anterior cruciate ligament (ACL), as well as extrusion from the tibial plateau. Similar to the evaluation of OA incidence, subgroups were delineated as subjects that remained healthy and those that developed meniscal injuries within four years, as determined by MOAKS grading. Damage to the posterior cruciate ligament was not considered due to insufficient number of grades.

To establish relationships between meniscus shape and future cartilage thickness trajectory, knees were repartitioned by PC score to either *Control-Associated Quartiles* or *OA Incidence-Associated Quartiles* for each statistically significant mode. Group analysis was similarly performed to compare average velocity between quartiles.

6.3 Results

6.3.1 Statistical Shape Modeling of the Meniscus

The primary ten modes of each meniscus are described in **Figure 6.3** and capture 90.17% and 87.71% of the total variance in the medial and lateral meniscus, respectively. More information regarding the compactness and cumulative geometric variance retained by these models can be found in **Figure 6.4**.

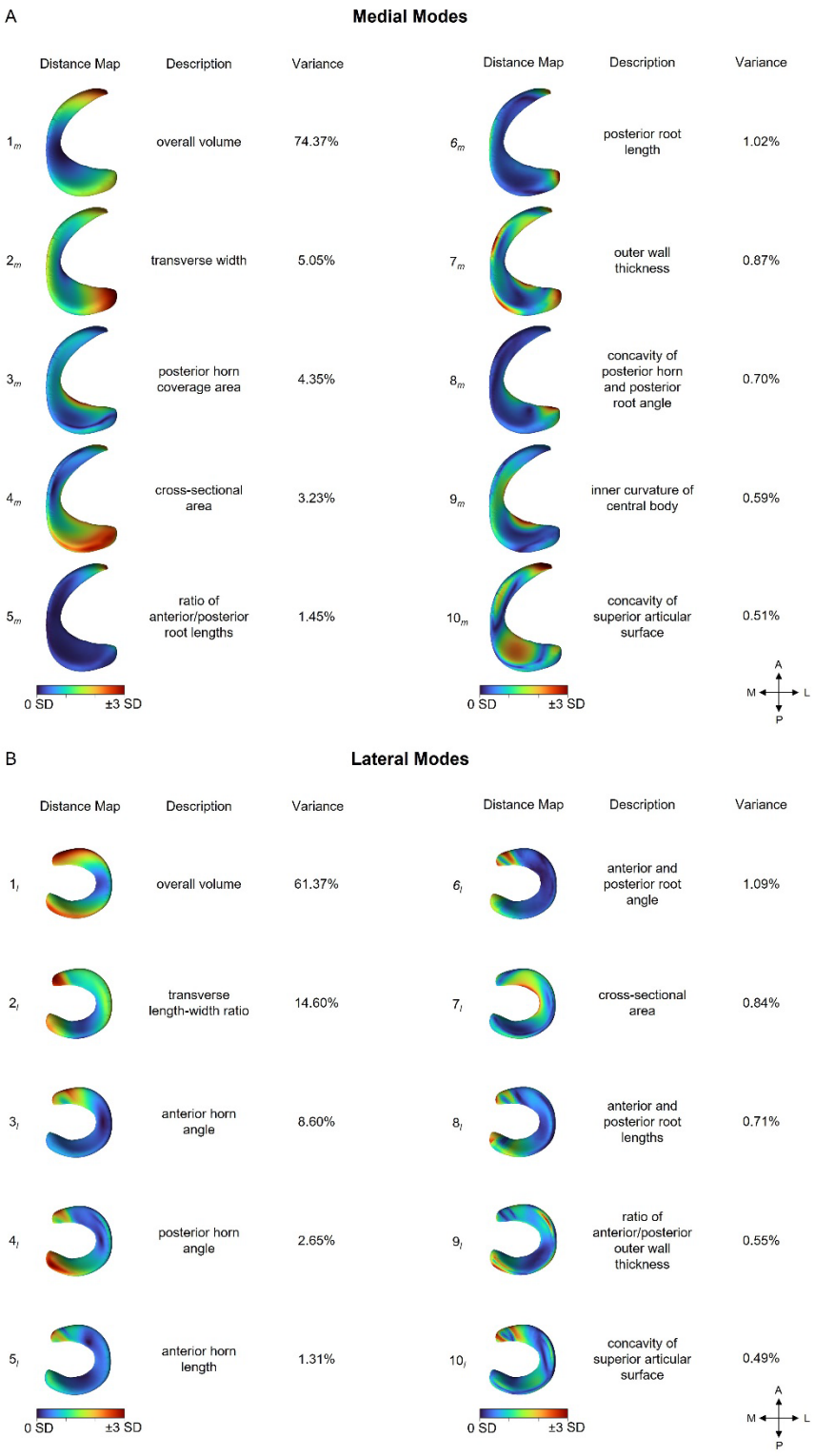

Figure 6.3 Modes of meniscus shape produced by PCA eigendecomposition of the geometric feature covariance. The visualization depicts the renderings of the average surface with color overlay of the Euclidean norm at ±3 standard deviations (SD). Modes are ordered by amount of variance captured by projection to each principal component.

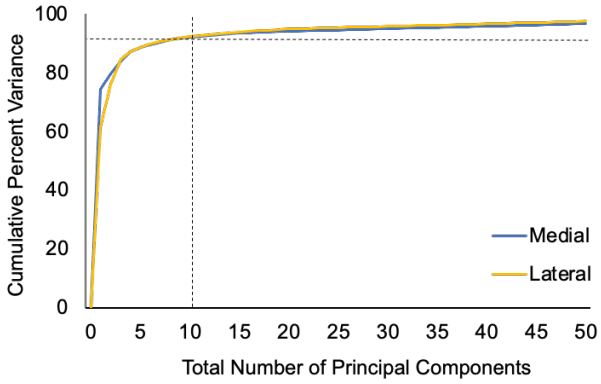

Figure 6.4 Cumulative percent variance captured by increasing number of PCs.

Demographic Analysis of Meniscus Shape

The association matrix, in **Figure 6.5A**, demonstrates cross-sectional relationships between meniscus shapes and demographics. Gender was the most prevalent category, with group differences between male and female in 17 of the observed 20 modes. The first mode, meniscal volume, was associated with all studied variables. Notably, shapes of the lateral meniscus were found to be more closely related to demographics than those of the medial meniscus.

Shape-Associated Development of Meniscal Damage

Figure 6.5B describes associations of meniscus shapes with future incidence of morphological damage. In the medial meniscus, posterior root tears and medial-side extrusion were the most predictable from shape information, with statistical dependence found in 8 and 7 out of 10 modes, respectively. Other relationships with meniscal tears, ACL tears, and anterior-side extrusion were less extensive. Patterns in the lateral meniscus were, again, more prevalent than in the medial meniscus. Mode 1, volume, was found to be associated with almost all forms of future damage.

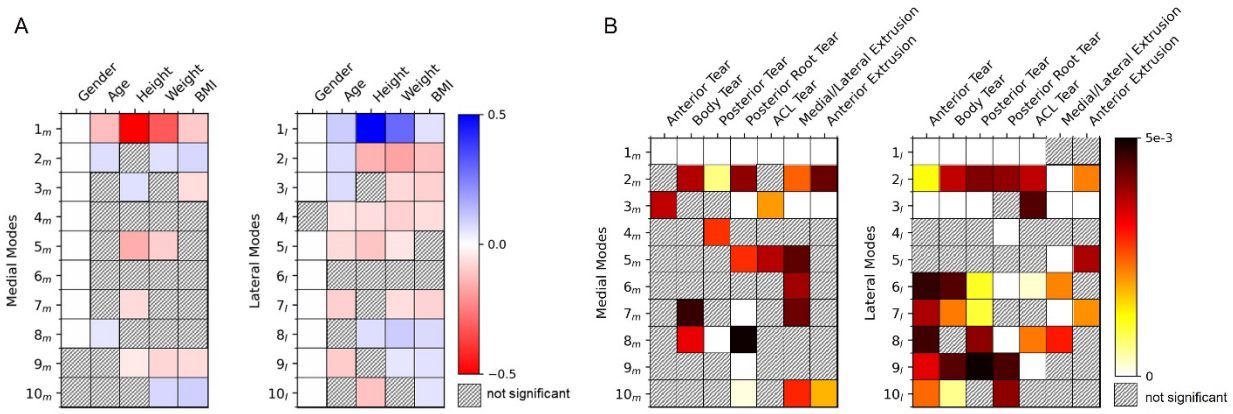

Figure 6.5 (A) Demographic analysis of meniscus shapes. Statistical significance (p < 0.005) was determined with ANCOVA for gender, and Pearson correlation for age, height, weight, and BMI. In the correlation analysis, the cell colors map to the R-value. (B) Prognostic analysis of meniscus shapes with presence of morphological imaging findings. Heatmap colors represent significant differences between menisci that develop damage within four years and those that remain normal, while grey cells depict non-statistical significance.

Meniscus Shapes as Risk Factor for Future OA

The *Control* group consisted of 2,778 subjects (1,536 female, age = 60.2 ± 9.2, BMI = 27.6 ± 4.5 kg/m^2) while the *OA Incidence* group contained 528 subjects (350 female, age = 60.4 ± 8.7, BMI = 29.1 ± 4.6 kg/m^2). Four of ten modes that describe medial meniscus shape and three of ten modes that describe lateral meniscus shape were significantly different between the *Control* and *OA Incidence* groups (**Figure 6.6**). Features of the medial meniscus that characterized the *OA Incidence* group were described as wider transverse diameter (mode 2_m: p < 0.001; % variance = 5.05%), longer anterior root accompanied by shorter posterior root (mode 5_m: p < 0.001; % variance = 1.45%), flaring of the outer wall (mode 7_m: p < 0.001; % variance = 0.87%), and increased concavity of the posterior horn and inward angling of the posterior root (mode 8_m: p < 0.001; %

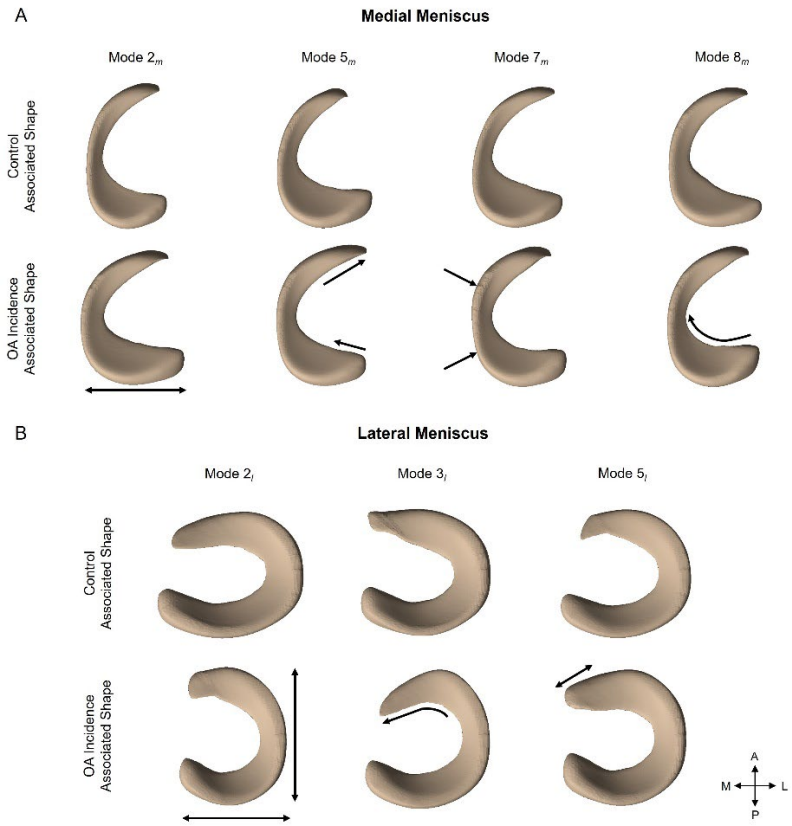

Figure 6.6 Statistically generated (A) medial and (B) lateral menisci representative of the *Control* and *OA Incidence* groups. The modes depicted, four in the medial side and three in the lateral side, were determined to be precursors to OA incidence. Arrows depict the qualitative shape interpretation that describe the *OA Incidence* group.

variance = 0.70%). While for the lateral meniscus, we observed the following in the *OA Incidence* group: larger transverse length-width ratio (mode 2_l: $p < 0.001$; % variance = 14.60%), inward angling (mode 3_l: $p < 0.001$; % variance = 8.60%) and increased length of the anterior horn (mode 5_l: $p < 0.001$; % variance = 1.31%).

Further investigation of the abovementioned shapes in conjunction with MOAKS grading of meniscal damage is detailed in **Table 6.2**. 703 participants from the *Control* group and 372 participants from the *OA Incidence* group were previously graded for MOAKS scoring of meniscal anterior, body, and posterior tears, posterior root tear, and anterior extrusion and included in this analysis. The effect size of each variable is described by the regression slope coefficient. To give physical meaning to the shape variable coefficients, the slopes are scaled by one SD, i.e. one-third of the deviation illustrated in **Figure 6.6**. In the medial meniscus, the addition of meniscal damage to the

regression did not change the statistical relationship of the shape variables. The coefficients of all four modes increased, indicating higher response to OA incidence after adjusting for tears and extrusion. Of the MOAKS variables, posterior tears and anterior extrusion were expectedly statistically associated with OA incidence. The effect size of

Table 6.2 Comparison of effect size of meniscal shapes with OA incidence

Independent Variable	Regression Slope Coefficient[c]	
	Shape Only	**Shape and Damage**
Medial Meniscus Shape		
Mode 2_m	0.0148*	0.0385*
Mode 5_m	0.0182**	0.0351*
Mode 7_m	0.0259**	0.0382*
Mode 8_m	0.0186**	0.0452*
Medial Meniscus Damage		
Anterior Tear	--	-0.0297
Body Tear	--	-0.0914
Posterior Tear	--	0.133*
Posterior Root Tear	--	-0.017
Anterior Extrusion	--	0.0985*
Lateral Meniscus Shape		
Mode 2_l	-0.00287*	0.00389
Mode 3_l	0.0449*	0.0368*
Mode 5_l	0.0299*	0.0267
Lateral Meniscus Damage		
Anterior Tear	--	0.153
Body Tear	--	-0.0724
Posterior Tear	--	-0.0063
Posterior Root Tear	--	-0.353
Anterior Extrusion	--	0.182*

*: $p < 0.05$
**: $p < 0.001$
[c]: Shape coefficients are scaled by one standard deviation

these forms of damage are equal to approximately 3 SD of the shape variables. In the lateral meniscus, anterior extrusions were associated with OA incidence, accounting for five times the average response of mode 3_l. Modes 2_l and 5_l were no longer statistically significant with the inclusion of meniscal damage variables.

6.3.2 Meniscus Shape-Associated Cartilage Thickness Trajectories

In the femur, knees of both the *Control* and *OA Incidence* groups were observed with overall cartilage thinning when spatially averaged: mean (SD) rate of −0.0026 (±0.0038) mm/year and −0.0035 (±0.0057) mm/year, respectively. 54.1% (±10.0%) of the cartilage surfaces in the *Control* velocity maps demonstrated thinning, as compared to 53.0% (±12.4%) of the surfaces in the *OA Incidence* maps. Similar patterns were seen in the tibial cartilage trajectories. Average velocities were −0.0044 (±0.0057) mm/year and −0.0067 (±0.0073) mm/year with thinning in 62.2% (12.0%) and 64.8% (12.3%) of the surfaces in the *Control* and *OA Incidence* groups, respectively.

Localized associations between meniscus shape and cartilage trajectory are shown in **Figure 6.7**. In all but one assessment (mode 8_m), the velocity differences were predominantly negative, implying lower overall velocities (*i.e.*, more rapid thinning or less rapid thickening) in the statistically significant regions of the *OA Incidence-Associated Quartiles* relative to those in the *Control-Associated Quartiles*. In general, regions of statistical difference in the femoral cartilage were more focal, whereas differences in the tibial cartilage were sparse and generally located around the cartilage periphery.

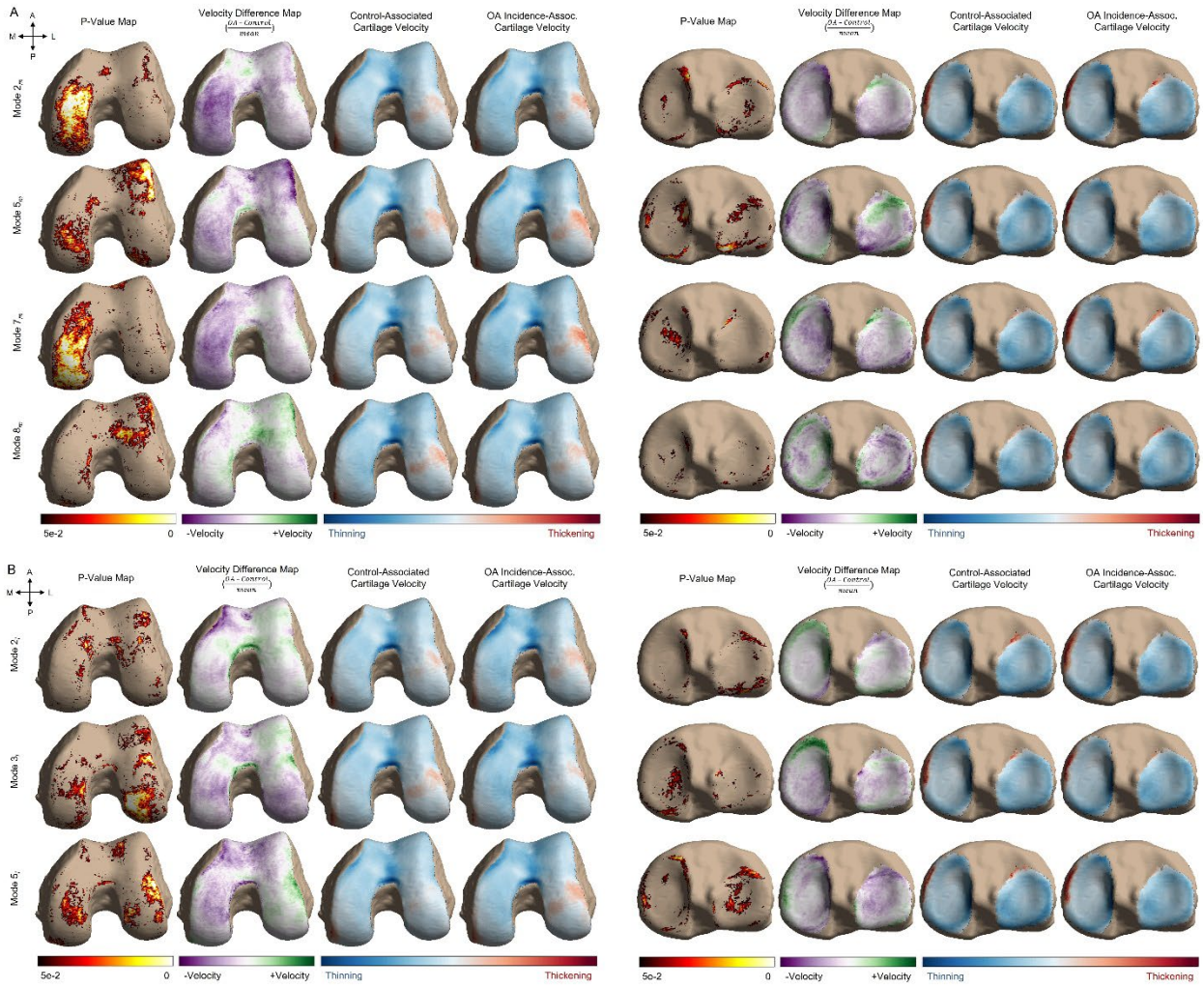

Figure 6.7 Cartilage thickness changes within the femoral and tibial compartments in relation to (A) medial and (B) lateral meniscus shapes. The significant focal cartilage thinning of the medial condyle in modes 2_m and 7_m may be attributed to decreased meniscal coverage. In mode 5_m, the increased length of the meniscal anterior horn may contribute to the thinning pattern in the opposite posterolateral tibia. The femoral velocity map of mode 8_m is the only studied feature in which the majority of significant points demonstrated decreased rates of cartilage thinning in the *Incidence* group. Mode 2_l, depicts a medial-lateral imbalance in the anterior aspect of the femoral cartilage. Mode 3_l, the inward angling of the meniscal anterior horn, may contribute to the significant thinning of femoral cartilage in the same area. Conversely, the increased length of the anterior horn in mode 5_l is associated with cartilage thickening in the lateral femoral cartilage and thinning in the lateral tibial compartment.

6.4 Discussion

In this study, we investigated the association of meniscus shape with future incidence of OA. Deep learning has enabled large-scale multi-tissue morphological evaluations. Combined with statistical shape modeling, an exploratory technique for parameterizing population-level shape features, human supervised feature engineering is no longer necessary to discover geometric characteristics of this nuanced tissue in relation to OA.

The characterization of meniscal geometry has been historically studied in the context of meniscus transplant. Several efforts have investigated non-invasive techniques for the optimization of meniscal replacement matching [241], [242], such as using demographic information as predictors for meniscus size. Modes 1_m and 1_l in this study, representative of meniscal volume, were found to have strong relationships with gender, age, height, weight, and BMI, relating well to past work [223], [225]. Our results describe gender as the most predictive of these demographics, which is consistent with a multivariate regression model developed by Van Thiel, *et al.* [241], that utilized similar variables in allograft sizing. Aside from gender, other basic demographics were statistically correlated to several shape features; however, the strengths of correlation were relatively weak with the exceptions of height and weight to meniscus size. Notwithstanding, these results serve as a preliminary benchmark for future analyses of meniscus shape stratification in demographic subpopulations.

Overall, the SSM produced anatomically consistent shape constructs. For example, structurally, the medial meniscus has more peripheral fixation, is less mobile than the lateral meniscus [243], and often moves as a single unit [244]. This is appreciable in that variations in the medial modes were generally radial and engages multiple sections of the

tissue, whereas those of the lateral meniscus were primarily found in the horns. Moreover, the medial posterior horn is the most anchored section of both menisci, and the most susceptible to damage [238]. In our results, shape variations in this area involved differences in thickness or cross-section, as opposed to length or position more commonly seen in the anterior or lateral horns. In relation to OA incidence, the shapes identified in this study are compatible with and expand upon past literature [223]–[225]. Features such as larger posterior wedge angle, smaller medial posterior wedge width [224], and bulging of the periphery found in OA populations [225], are related to modes identified here as OA precursors (modes 7_m and 8_m). However, these relationships should be evaluated with consideration to interactions between shape and tissue damage or injury. The multivariate sub-analysis with shape and MOAKS grading in **Table 6.2** found consistent associations with OA incidence in the medial meniscus, but two of three relationships of the lateral meniscus were no longer statistically significant, suggesting tears and extrusions playing confounding roles. It is important to note that further longitudinal analysis using this framework can help identify how meniscus shape progresses over time, both in the presence and absence of OA.

A key strength of this approach is the joint assessment of meniscal shape with future cartilage thickness changes to reveal localized patterns of knee joint degeneration. The seven shapes identified as OA risk factors each presented with unique cartilage trajectories. Group differences in the femoral compartments were more prevalent than those in the tibia, likely due to the complexity of meniscal shape being primarily in the concave meniscal-femoral interface as opposed to the flat, fixated meniscal-tibial surface.

Differences in cartilage thickness changes due to medial meniscus shapes were more prevalent than lateral meniscus shapes. Modes 2_m and 7_m both presented with large, central regions of increased rates of thinning in the medial femoral condyle of the *OA Incidence-Associated Quartile*, despite describing distinct shape features (wider transverse diameter and flaring of outer wall, respectively). Considering that both modes were also identified as predictors of extrusion, we suspect that these modes are linked to decreased central coverage in the medial compartment and progressive cartilage degeneration. Differences in mode 5_m were more divergent, with moderately increased rates of thinning in the medial compartment and dramatically increased rates in the opposite lateral trochlea. Similarly, we observed increased thinning near the medial tibial eminence accompanied by decreased thinning in the anterolateral plateau. The meniscal shape, interpreted as an elongation of the anterior horn and perhaps influenced by the presence or prominence of the transverse ligament, may suggest rotational imbalance that causes pressure points along the medial wall of the intercondylar notch and lateral trochlea. Lastly, mode 8_m also presented with focal differences in the lateral trochlea, yet the cartilage of the *OA Incidence-Associated Quartile* thinned at a slower rate than the *Control-Associated Quartile*. This mode may be more closely related to non-cartilage subtypes or OA of the patellofemoral joint.

Relative to medial meniscus shape-related cartilage trajectories, those associated with lateral meniscus shapes were sparser. Regions of significant differences were located on both medial and lateral sides in the femoral, trochlear, and tibial compartments. Notably, medial regions tend to describe increased rates of thinning in the *OA Incidence-Associated Quartile* whereas lateral regions presented with decreased rates. This

suggests that lateral meniscus shape may play a secondary role in this population with predominantly medial compartment OA.

An advantage of this approach is that it is quantitative and automated, however this also introduces some limitations. While this methodology is capable of evaluating meniscus, cartilage, and bone collectively, it lacks true localization of the overall joint structure. Secondly, discoid meniscus, an abnormal and congenital development involving morphological and structural deformations, was not evaluated in our study due to low and varied incidence rates ranging from 0.06-17% [245]. It is evident from the cartilage analysis that the trochlea, and possibly the patellar cartilage, are impacted by meniscus shape. In this work, we constrained our scope to tibiofemoral OA, yet it would be informative to evaluate patellofemoral OA in future studies.

In summary, this data-driven method presented a general foundation of common variations of meniscus geometry, broadened the description of meniscus characteristics that are associated with the onset of OA, and also discovered novel shape features that have yet to be investigated in the context of OA risk. Furthermore, localized longitudinal changes in cartilage thickness were associated with each meniscus shape, contributing additional granularity to prognosis of cartilage degeneration. This work describes a quantitative approach toward integrating meniscus morphology into the assessment of knee OA as a whole joint disease.

Acknowledgements

This work was funded by the National Institutes of Health and National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant number: R00AR070902). The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259;

N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

Chapter 7

Summary and Future Directions

There has been rapid advancement in the application of DL techniques in medical imaging. When joined with recent growth in the rate of data collection and the explosion in the capacity of computing power, there is immense potential for computer vision to push the boundaries of our knowledge of disease.

MSK imaging, in particular, is an apt candidate for these applications. We demonstrated the utility of fully convolutional encoder-decoders for tissue segmentation in lumbar spine MRI. Through our experimental process, we present guidelines for adapting techniques proven useful in natural images to clinical MRI, overcoming challenges such as variations in acquisition protocols, highly anisotropic voxels, and small sample sizes. We employed these models in subject-specific biomechanical models, showing that machine-derived segmentations were highly correlated with those generated with human segmentations.

DL-based tissue segmentation served as the foundation of this dissertation. Upon it, we constructed pipelines for anomaly detection, geometric analysis, and explored clinical applications.

In the Modic mapping tool, we presented the use of segmentation for lesion detection. This tool was designed for clinical utility, producing intermediate results at multiple stages for human interpretation. Though not intended for use as a standalone diagnostic tool, we found that the introduction of Modic maps as computer-generated aid was beneficial for improving agreement between radiologists-in-training and attendings. Notably, this technique promoted heterogeneous assessment of MCs. Capturing tissue features during the transitional phase is challenging for the human eye. We hope that the granular nature of the voxel-wise mapping will encourage more attention to mixed-type MCs and their relationship to pain.

Beyond clinical imaging, tissue segmentation has increased our capacity to assess advanced imaging techniques. Imaging sequences like compositional MRI and high-resolution Cube would traditionally have high labor costs for analysis. With the application of DL segmentation, multiple tissues spanning hundreds of slices and numerous TSL/TE acquisitions in a single MR exam can be expeditiously delineated. Through automation of tissue extraction, we are able to efficiently perform multiparametric analysis, leveraging information from an assortment of MR sequences. In our study of knee morphology in young athletes, we determined that basketball players had elevated $T_{1\rho}$ and T_2 in the posteromedial and posterolateral femoral cartilage and deep layer of patellar cartilage with decreased values in the superficial layer of patellar cartilage. These patterns were not detected in the classical ROI-based analysis, demonstrating the utility of the spatially

sensitive analysis of VBR. Furthermore, we found three bone shape variations between basketball players with high-knee impact and swimmers with low knee-impact. These relationships characterized the knees of young athletes with imaging alone and provide compelling direction in terms of understanding how to minimize degenerative changes while keeping these young athletes at the top of their game.

In the final chapter of this work, we applied the earlier techniques in a population-level effort to characterize meniscus shape. Leveraging DL techniques and experimental design, tissues were segmented from 4,790 subjects of the OAI dataset. Adopting statistical shape modeling used for bone shape analysis, a meniscus shape model was built, forming a data-driven PCA-space of general variations in meniscal geometry. Our statistical analysis found seven meniscal shapes that were more prevalent in populations that eventually developed knee OA. The key distinction to this methodology is the incorporation of cartilage and bone geometry. We investigated the relationship between meniscus shape and longitudinal changes in cartilage by encoding the trajectory of cartilage thickness to points on the subchondral bone. Knee OA is a whole joint disease, and this work contributes a quantitative approach to the prognosis of cartilage degeneration in relation to its neighboring tissues.

Returning to the primary objective of improving outcomes of LBP and knee OA, there are two paths of future research that I hope would unfold from these works.

In the clinical application of LBP, the models we developed were intended for use in the BACPAC consortium (https://heal.nih.gov/research/clinical-research/back-pain), a large, multicenter effort to develop a state-of-the-art model for cLBP. Though we have demonstrated excellent model performance using images from our institution, we have

yet to perform validation with images from other BACPAC sites. Despite efforts to harmonize imaging protocols across the consortium, we expect domain shifts with external data. Further model fine-tuning may be necessary to improve generalizability. Self supervised learning may be a fitting contender for this purpose. RadImageNet [246] is a medical imaging dataset containing 1.35 million MRI, computed tomography, and ultrasound images. Without the need for annotation, pretraining with self supervised techniques may induce more robust learning and model convergence at more generalizable local minima.

With respect to knee OA, there is exciting work currently being pursued to holistically characterize the knee joint from a single MRI exam. In this effort, we generated a 100-dimensional PC space, composed of various OA biomarkers that were discussed in this dissertation, spanning from tissue geometry to cartilage compositional mapping. Validated on the OAI dataset, this multiparametric characterization of the knee will allow for large-scale analyses of the relationships between tissue composition, morphology, and OA.

In summary, computer vision and deep learning have enabled tissue morphology to be assessed in a quantitative and exploratory manner. Feature extraction no longer requires human engineering but can rather be performed through data-driven pattern recognition. Through the methods presented here, we developed robust models and assessed the interaction between various tissues of interest, bringing us closer to evaluating these heterogeneous diseases with a multifaceted lens.

References

[1] “BMUS: The Burden of Musculoskeletal Diseases in the United States,” *BMUS: The Burden of Musculoskeletal Diseases in the United States*. https://www.boneandjointburden.org/ (accessed Apr. 28, 2023).

[2] T. Vos *et al.*, “Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015,” *The Lancet*, vol. 388, no. 10053, pp. 1545–1602, Oct. 2016, doi: 10.1016/S0140-6736(16)31678-6.

[3] J. Hong, C. Reed, D. Novick, and M. Happich, “Costs associated with treatment of chronic low back pain: an analysis of the UK General Practice Research Database,” *Spine (Phila Pa 1976)*, vol. 38, no. 1, pp. 75–82, Jan. 2013, doi: 10.1097/BRS.0b013e318276450f.

[4] B. F. Walker, R. Muller, and W. D. Grant, “Low back pain in Australian adults: the economic burden,” *Asia Pac J Public Health*, vol. 15, no. 2, pp. 79–87, 2003, doi: 10.1177/101053950301500202.

[5] M. L. Ferreira, G. Machado, J. Latimer, C. Maher, P. H. Ferreira, and R. J. Smeets, “Factors defining care-seeking in low back pain--a meta-analysis of population based surveys,” *Eur J Pain*, vol. 14, no. 7, p. 747.e1–7, Aug. 2010, doi: 10.1016/j.ejpain.2009.11.005.

[6] GBD 2013 DALYs and HALE Collaborators *et al.*, “Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological

transition," *Lancet*, vol. 386, no. 10009, pp. 2145–2191, Nov. 2015, doi: 10.1016/S0140-6736(15)61340-X.

[7] D. J. Schofield, E. J. Callander, R. N. Shrestha, M. E. Passey, S. J. Kelly, and R. Percival, "Back problems, comorbidities, and their association with wealth," *Spine J*, vol. 15, no. 1, pp. 34–41, Jan. 2015, doi: 10.1016/j.spinee.2014.06.018.

[8] R. Froud *et al.*, "A systematic review and meta-synthesis of the impact of low back pain on people's lives," *BMC Musculoskelet Disord*, vol. 15, p. 50, Feb. 2014, doi: 10.1186/1471-2474-15-50.

[9] R. Shiri, J. Karppinen, P. Leino-Arjas, S. Solovieva, and E. Viikari-Juntura, "The association between smoking and low back pain: a meta-analysis," *Am J Med*, vol. 123, no. 1, p. 87.e7–35, Jan. 2010, doi: 10.1016/j.amjmed.2009.05.028.

[10] T.-T. Zhang, Z. Liu, Y.-L. Liu, J.-J. Zhao, D.-W. Liu, and Q.-B. Tian, "Obesity as a Risk Factor for Low Back Pain: A Meta-Analysis," *Clin Spine Surg*, vol. 31, no. 1, pp. 22–27, Feb. 2018, doi: 10.1097/BSD.0000000000000468.

[11] R. Shiri and K. Falah-Hassani, "Does leisure time physical activity protect against low back pain? Systematic review and meta-analysis of 36 prospective cohort studies," *Br J Sports Med*, vol. 51, no. 19, pp. 1410–1418, Oct. 2017, doi: 10.1136/bjsports-2016-097352.

[12] P. H. Ferreira, P. Beckenkamp, C. G. Maher, J. L. Hopper, and M. L. Ferreira, "Nature or nurture in low back pain? Results of a systematic review of studies based on twin samples," *Eur J Pain*, vol. 17, no. 7, pp. 957–971, Aug. 2013, doi: 10.1002/j.1532-2149.2012.00277.x.

[13] D. Steffens *et al.*, “What triggers an episode of acute low back pain? A case-crossover study,” *Arthritis Care Res (Hoboken)*, vol. 67, no. 3, pp. 403–410, Mar. 2015, doi: 10.1002/acr.22533.

[14] P. do Carmo Silva Parreira *et al.*, “Can patients identify what triggers their back pain? Secondary analysis of a case-crossover study,” *Pain*, vol. 156, no. 10, pp. 1913–1919, Oct. 2015, doi: 10.1097/j.pain.0000000000000252.

[15] B. R. Waterman, P. J. Belmont, and A. J. Schoenfeld, “Low back pain in the United States: incidence and risk factors for presentation in the emergency setting,” *Spine J*, vol. 12, no. 1, pp. 63–70, Jan. 2012, doi: 10.1016/j.spinee.2011.09.002.

[16] J. Hartvigsen *et al.*, “What low back pain is and why we need to pay attention,” *The Lancet*, vol. 391, no. 10137, pp. 2356–2367, Jun. 2018, doi: 10.1016/S0140-6736(18)30480-X.

[17] C. Maher, M. Underwood, and R. Buchbinder, “Non-specific low back pain,” *The Lancet*, vol. 389, no. 10070, pp. 736–747, Feb. 2017, doi: 10.1016/S0140-6736(16)30970-9.

[18] E. I. T. de Schepper *et al.*, “The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features,” *Spine (Phila Pa 1976)*, vol. 35, no. 5, pp. 531–536, Mar. 2010, doi: 10.1097/BRS.0b013e3181aa5b33.

[19] B.-G. Peng, “Pathophysiology, diagnosis, and treatment of discogenic low back pain,” *World J Orthop*, vol. 4, no. 2, pp. 42–52, Apr. 2013, doi: 10.5312/wjo.v4.i2.42.

[20] K. Fujii *et al.*, “Discogenic Back Pain: Literature Review of Definition, Diagnosis, and Treatment,” *JBMR Plus*, vol. 3, no. 5, p. e10180, Mar. 2019, doi: 10.1002/jbm4.10180.

[21] N. A. Farshad-Amacker, A. Hughes, R. J. Herzog, B. Seifert, and M. Farshad, “The intervertebral disc, the endplates and the vertebral bone marrow as a unit in the process of degeneration,” *Eur Radiol*, vol. 27, no. 6, pp. 2507–2520, Jun. 2017, doi: 10.1007/s00330-016-4584-z.

[22] R. J. Moore, “The vertebral endplate: disc degeneration, disc regeneration,” *Eur Spine J*, vol. 15, no. Suppl 3, pp. 333–337, Aug. 2006, doi: 10.1007/s00586-006-0170-4.

[23] J. C. Lotz, A. J. Fields, and E. C. Liebenberg, “The Role of the Vertebral End Plate in Low Back Pain,” *Global Spine J*, vol. 3, no. 3, pp. 153–164, Jun. 2013, doi: 10.1055/s-0033-1347298.

[24] P. Suri *et al.*, “Does lumbar spinal degeneration begin with the anterior structures? A study of the observed epidemiology in a community-based population,” *BMC Musculoskelet Disord*, vol. 12, p. 202, Sep. 2011, doi: 10.1186/1471-2474-12-202.

[25] A. Fujiwara *et al.*, “The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study,” *Eur Spine J*, vol. 8, no. 5, pp. 396–401, 1999, doi: 10.1007/s005860050193.

[26] M. J. Hancock *et al.*, “Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain,” *Eur Spine J*, vol. 16, no. 10, pp. 1539–1550, Oct. 2007, doi: 10.1007/s00586-007-0391-1.

[27] S.-H. Lee, D. W. Son, J.-S. Lee, S.-K. Sung, S. W. Lee, and G. S. Song, “Relationship Between Endplate Defects, Modic Change, Facet Joint Degeneration, and Disc Degeneration of Cervical Spine,” *Neurospine*, vol. 17, no. 2, pp. 443–452, Jun. 2020, doi: 10.14245/ns.2040076.038.

[28] D. Steffens, M. J. Hancock, C. G. Maher, C. Williams, T. S. Jensen, and J. Latimer, “Does magnetic resonance imaging predict future low back pain? A systematic review,” *European Journal of Pain*, vol. 18, no. 6, pp. 755–765, 2014, doi: 10.1002/j.1532-2149.2013.00427.x.

[29] W. Brinjikji *et al.*, “Systematic literature review of imaging features of spinal degeneration in asymptomatic populations,” *AJNR Am J Neuroradiol*, vol. 36, no. 4, pp. 811–816, Apr. 2015, doi: 10.3174/ajnr.A4173.

[30] R. Chou, A. Qaseem, D. K. Owens, P. Shekelle, and Clinical Guidelines Committee of the American College of Physicians, “Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians,” *Ann Intern Med*, vol. 154, no. 3, pp. 181–189, Feb. 2011, doi: 10.7326/0003-4819-154-3-201102010-00008.

[31] M. J. Hancock, P. Kjaer, P. Kent, R. K. Jensen, and T. S. Jensen, “Is the Number of Different MRI Findings More Strongly Associated With Low Back Pain Than Single MRI Findings?,” *Spine*, vol. 42, no. 17, pp. 1283–1288, Sep. 2017, doi: 10.1097/BRS.0000000000002102.

[32] R. Sihawong, P. Janwantanakul, and W. Jiamjarasrangsi, “A prospective, cluster-randomized controlled trial of exercise program to prevent low back pain in office

workers," *Eur Spine J*, vol. 23, no. 4, pp. 786–793, Apr. 2014, doi: 10.1007/s00586-014-3212-3.

[33] D. Steffens *et al.*, "Prevention of Low Back Pain: A Systematic Review and Meta-analysis," *JAMA Intern Med*, vol. 176, no. 2, pp. 199–208, Feb. 2016, doi: 10.1001/jamainternmed.2015.7431.

[34] G. Varrassi, B. Moretti, M. C. Pace, P. Evangelista, and G. Iolascon, "Common Clinical Practice for Low Back Pain Treatment: A Modified Delphi Study," *Pain Ther*, vol. 10, no. 1, pp. 589–604, Jun. 2021, doi: 10.1007/s40122-021-00249-w.

[35] F. W. Roemer *et al.*, "Structural phenotypes of knee osteoarthritis: potential clinical and research relevance," *Skeletal Radiol*, Sep. 2022, doi: 10.1007/s00256-022-04191-6.

[36] S. Safiri *et al.*, "Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017," *Annals of the Rheumatic Diseases*, vol. 79, no. 6, pp. 819–828, Jun. 2020, doi: 10.1136/annrheumdis-2019-216515.

[37] A. Cui, H. Li, D. Wang, J. Zhong, Y. Chen, and H. Lu, "Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies," *eClinicalMedicine*, vol. 29, Dec. 2020, doi: 10.1016/j.eclinm.2020.100587.

[38] D. J. Hunter and S. Bierma-Zeinstra, "Osteoarthritis," *The Lancet*, vol. 393, no. 10182, pp. 1745–1759, Apr. 2019, doi: 10.1016/S0140-6736(19)30417-9.

[39] H. Long *et al.*, "Prevalence Trends of Site-Specific Osteoarthritis From 1990 to 2019: Findings From the Global Burden of Disease Study 2019," *Arthritis & Rheumatology*, vol. 74, no. 7, pp. 1172–1183, 2022, doi: 10.1002/art.42089.

[40] J. N. Katz, K. R. Arant, and R. F. Loeser, “Diagnosis and Treatment of Hip and Knee Osteoarthritis: A Review,” *JAMA*, vol. 325, no. 6, pp. 568–578, Feb. 2021, doi: 10.1001/jama.2020.22171.

[41] Y. Zhang and J. M. Jordan, “Epidemiology of Osteoarthritis,” *Clin Geriatr Med*, vol. 26, no. 3, pp. 355–369, Aug. 2010, doi: 10.1016/j.cger.2010.03.001.

[42] M. A. Davis, W. H. Ettinger, J. M. Neuhaus, J. D. Barclay, and M. R. Segal, “Correlates of knee pain among US adults with and without radiographic knee osteoarthritis,” *J Rheumatol*, vol. 19, no. 12, pp. 1943–1949, Dec. 1992.

[43] T. Neogi *et al.*, “Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies,” *BMJ*, vol. 339, p. b2844, Aug. 2009, doi: 10.1136/bmj.b2844.

[44] J. Bedson and P. R. Croft, “The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature,” *BMC Musculoskeletal Disorders*, vol. 9, no. 1, p. 116, Sep. 2008, doi: 10.1186/1471-2474-9-116.

[45] S. Glyn-Jones *et al.*, “Osteoarthritis,” *The Lancet*, vol. 386, no. 9991, pp. 376–387, Jul. 2015, doi: 10.1016/S0140-6736(14)60802-3.

[46] M. Maldonado and J. Nam, “The Role of Changes in Extracellular Matrix of Cartilage in the Presence of Inflammation on the Pathology of Osteoarthritis,” *Biomed Res Int*, vol. 2013, p. 284873, 2013, doi: 10.1155/2013/284873.

[47] J. Runhaar *et al.*, “Prevention of knee osteoarthritis in overweight females: the first preventive randomized controlled trial in osteoarthritis,” *Am J Med*, vol. 128, no. 8, pp. 888-895.e4, Aug. 2015, doi: 10.1016/j.amjmed.2015.03.006.

[48] M. Fransen, S. McConnell, A. R. Harmer, M. Van der Esch, M. Simic, and K. L. Bennell, “Exercise for osteoarthritis of the knee: a Cochrane systematic review,” *Br J Sports Med*, vol. 49, no. 24, pp. 1554–1557, Dec. 2015, doi: 10.1136/bjsports-2015-095424.

[49] M. Hall, B. Castelein, R. Wittoek, P. Calders, and A. Van Ginckel, “Diet-induced weight loss alone or combined with exercise in overweight or obese people with knee osteoarthritis: A systematic review and meta-analysis,” *Semin Arthritis Rheum*, vol. 48, no. 5, pp. 765–777, Apr. 2019, doi: 10.1016/j.semarthrit.2018.06.005.

[50] A. E. Nelson, K. D. Allen, Y. M. Golightly, A. P. Goode, and J. M. Jordan, “A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative,” *Semin Arthritis Rheum*, vol. 43, no. 6, pp. 701–712, Jun. 2014, doi: 10.1016/j.semarthrit.2013.11.012.

[51] P. Jüni *et al.*, “Intra-articular corticosteroid for knee osteoarthritis,” *Cochrane Database Syst Rev*, vol. 2015, no. 10, p. CD005328, Oct. 2015, doi: 10.1002/14651858.CD005328.pub3.

[52] J. B. Thorlund, C. B. Juhl, E. M. Roos, and L. S. Lohmander, “Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms,” *BMJ*, vol. 350, p. h2747, Jun. 2015, doi: 10.1136/bmj.h2747.

[53] R. Brignardello-Petersen *et al.*, “Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review,” *BMJ Open*, vol. 7, no. 5, p. e016114, May 2017, doi: 10.1136/bmjopen-2017-016114.

[54] M. M. Dowsey, M. Nikpour, P. Dieppe, and P. F. Choong, “Associations between pre-operative radiographic changes and outcomes after total hip replacement for osteoarthritis,” *Osteoarthritis and Cartilage*, vol. 22, pp. S174–S175, Apr. 2014, doi: 10.1016/j.joca.2014.02.329.

[55] V. Wylde, S. Hewlett, I. D. Learmonth, and P. Dieppe, “Persistent pain after joint replacement: prevalence, sensory qualities, and postoperative determinants,” *Pain*, vol. 152, no. 3, pp. 566–572, Mar. 2011, doi: 10.1016/j.pain.2010.11.023.

[56] S. T. Skou *et al.*, “A Randomized, Controlled Trial of Total Knee Replacement,” *New England Journal of Medicine*, vol. 373, no. 17, pp. 1597–1606, Oct. 2015, doi: 10.1056/NEJMoa1505467.

[57] S. Bosshard, “Non-invasive assessment of sensory and nociceptive processing in mice using functional MRI,” Jan. 2011, doi: 10.5167/uzh-58943.

[58] L. Loevner and T. Booth, “ACR – ASNR – SCBT-MR PRACTICE PARAMETER FOR THE PERFORMANCE OF MAGNETIC RESONANCE IMAGING (MRI) OF THE ADULT SPINE,” 2014. Accessed: May 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/ACR-%E2%80%93-ASNR-%E2%80%93-SCBT-MR-PRACTICE-PARAMETER-FOR-THE-OF-Loevner-Booth/15404804a4cdf40d78f33ea1132bc4f3b5e4ba27

[59] V. Romeo *et al.*, “Clinical impact of coronal-STIR sequence in a routine lumbar spine MR imaging protocol to investigate low back pain,” *Medicine*, vol. 97, no. 23, p. e10789, Jun. 2018, doi: 10.1097/MD.0000000000010789.

[60] C. G. Peterfy, G. Gold, F. Eckstein, F. Cicuttini, B. Dardzinski, and R. Stevens, “MRI protocols for whole-organ assessment of the knee in osteoarthritis,” *Osteoarthritis and Cartilage*, vol. 14, pp. 95–111, Jan. 2006, doi: 10.1016/j.joca.2006.02.029.

[61] C. G. Peterfy *et al.*, “Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis,” *Osteoarthritis and Cartilage*, vol. 12, no. 3, pp. 177–190, Mar. 2004, doi: 10.1016/j.joca.2003.11.003.

[62] D. J. Hunter *et al.*, “Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score),” *Osteoarthritis and Cartilage*, vol. 19, no. 8, pp. 990–1002, Aug. 2011, doi: 10.1016/j.joca.2011.05.004.

[63] D. Hayashi, F. W. Roemer, and A. Guermazi, “Imaging of osteoarthritis—recent research developments and future perspective,” *Br J Radiol*, vol. 91, no. 1085, p. 20170349, May 2018, doi: 10.1259/bjr.20170349.

[64] C. W. Pfirrmann, A. Metzdorf, M. Zanetti, J. Hodler, and N. Boos, “Magnetic resonance classification of lumbar intervertebral disc degeneration,” *Spine (Phila Pa 1976)*, vol. 26, no. 17, pp. 1873–1878, Sep. 2001, doi: 10.1097/00007632-200109010-00011.

[65] M.-V. Corniola *et al.*, “Correlation of pain, functional impairment, and health-related quality of life with radiological grading scales of lumbar degenerative disc disease,” *Acta Neurochir (Wien)*, vol. 158, no. 3, pp. 499–505, Mar. 2016, doi: 10.1007/s00701-015-2700-5.

[66] J. Zheng and C. Shen, “Quantitative Relationship between the Degree of Lumbar Disc Degeneration and Intervertebral Disc Height in Patients with Low Back Pain,”

Contrast Media Mol Imaging, vol. 2022, p. 5960317, Jul. 2022, doi: 10.1155/2022/5960317.

[67] F. W. Roemer *et al.*, “State of the Art: Imaging of Osteoarthritis—Revisited 2020,” *Radiology*, vol. 296, no. 1, pp. 5–21, Jul. 2020, doi: 10.1148/radiol.2020192498.

[68] T. J. Mosher *et al.*, “Knee Articular Cartilage Damage in Osteoarthritis: Analysis of MR Image Biomarker Reproducibility in ACRIN-PA 4001 Multicenter Trial,” *Radiology*, vol. 258, no. 3, pp. 832–842, Mar. 2011, doi: 10.1148/radiol.10101174.

[69] S. T. Soellner, A. Goldmann, D. Muelheims, G. H. Welsch, and M. L. Pachowsky, “Intraoperative validation of quantitative T2 mapping in patients with articular cartilage lesions of the knee,” *Osteoarthritis Cartilage*, vol. 25, no. 11, pp. 1841–1849, Nov. 2017, doi: 10.1016/j.joca.2017.07.021.

[70] V. Casula *et al.*, “Validation and optimization of adiabatic T1ρ and T2ρ for quantitative imaging of articular cartilage at 3 T,” *Magn Reson Med*, vol. 77, no. 3, pp. 1265–1275, Mar. 2017, doi: 10.1002/mrm.26183.

[71] R. R. Regatte, S. V. S. Akella, A. Borthakur, J. B. Kneeland, and R. Reddy, “In vivo proton MR three-dimensional T1rho mapping of human articular cartilage: initial experience,” *Radiology*, vol. 229, no. 1, pp. 269–274, Oct. 2003, doi: 10.1148/radiol.2291021041.

[72] L. Wang and R. R. Regatte, “Quantitative Mapping of Human Cartilage at 3.0T,” *Acad Radiol*, vol. 21, no. 4, pp. 463–471, Apr. 2014, doi: 10.1016/j.acra.2013.12.010.

[73] D. E. T. Shepherd and B. B. Seedhom, “Thickness of human articular cartilage in joints of the lower limb,” *Annals of the Rheumatic Diseases*, vol. 58, no. 1, pp. 27–34, Jan. 1999, doi: 10.1136/ard.58.1.27.

[74] Y. Wang, M. C. Battié, S. K. Boyd, and T. Videman, “The osseous endplates in lumbar vertebrae: thickness, bone mineral density and their associations with age and disk degeneration,” *Bone*, vol. 48, no. 4, pp. 804–809, Apr. 2011, doi: 10.1016/j.bone.2010.12.005.

[75] T. Ai, W. Zhang, N. K. Priddy, and X. Li, “Diagnostic performance of CUBE MRI sequences of the knee compared with conventional MRI,” *Clin Radiol*, vol. 67, no. 12, pp. e58-63, Dec. 2012, doi: 10.1016/j.crad.2012.07.020.

[76] H. J. Park *et al.*, “Three-dimensional isotropic T2-weighted fast spin-echo (VISTA) knee MRI at 3.0 T in the evaluation of the anterior cruciate ligament injury with additional views: comparison with two-dimensional fast spin-echo T2-weighted sequences,” *Acta Radiol*, vol. 57, no. 11, pp. 1372–1379, Nov. 2016, doi: 10.1177/0284185114568048.

[77] P. Dodin, J.-P. Pelletier, J. Martel-Pelletier, and F. Abram, “Automatic human knee cartilage segmentation from 3D magnetic resonance images,” *IEEE Trans Biomed Eng*, vol. 57, no. 11, Nov. 2010, doi: 10.1109/TBME.2010.2058112.

[78] J. Carballido-Gamio, J. Bauer, K.-Y. Lee, S. Krause, and S. Majumdar, “Combined image processing techniques for characterization of MRI cartilage of the knee,” *Conf Proc IEEE Eng Med Biol Soc*, vol. 2005, pp. 3043–3046, 2005, doi: 10.1109/IEMBS.2005.1617116.

[79] J. Duryea *et al.*, “Local area cartilage segmentation: a semiautomated novel method of measuring cartilage loss in knee osteoarthritis,” *Arthritis Care Res (Hoboken)*, vol. 66, no. 10, pp. 1560–1565, Oct. 2014, doi: 10.1002/acr.22332.

[80] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int J Comput Vis*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[81] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2012. Accessed: May 07, 2023. [Online]. Available: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

[82] H. Afan *et al.*, “Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques,” *Engineering Applications of Computational Fluid Mechanics*, vol. 15, pp. 1420–1439, Sep. 2021, doi: 10.1080/19942060.2021.1974093.

[83] K. Fukushima, “Cognitron: A self-organizing multilayered neural network,” *Biol. Cybernetics*, vol. 20, no. 3, pp. 121–136, Sep. 1975, doi: 10.1007/BF00342633.

[84] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *arXiv:2010.11929 [cs]*, Jun. 2021, Accessed: Feb. 25, 2022. [Online]. Available: http://arxiv.org/abs/2010.11929

[85] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *arXiv:1505.04597 [cs]*, May 2015, Accessed: Mar. 14, 2022. [Online]. Available: http://arxiv.org/abs/1505.04597

[86] J. McGlinchy, B. Johnson, B. Muller, M. Joseph, and J. Diaz, “Application of UNet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment from High Resolution Satellite Imagery,” in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2019, pp. 3915–3918. doi: 10.1109/IGARSS.2019.8900453.

[87] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, “U-Net and its variants for medical image segmentation: theory and applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.

[88] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.” arXiv, Jun. 21, 2016. doi: 10.48550/arXiv.1606.06650.

[89] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nat Methods*, vol. 18, no. 2, Art. no. 2, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[90] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” *arXiv:1606.04797 [cs]*, Jun. 2016, Accessed: Feb. 05, 2021. [Online]. Available: http://arxiv.org/abs/1606.04797

[91] H. Cao *et al.*, “Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation.” arXiv, May 12, 2021. doi: 10.48550/arXiv.2105.05537.

[92] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[93] Z. Liu, Z. Xu, J. Jin, Z. Shen, and T. Darrell, “Dropout Reduces Underfitting.” arXiv, Mar. 02, 2023. doi: 10.48550/arXiv.2303.01500.

[94] B. J. Kim, H. Choi, H. Jang, D. Lee, and S. W. Kim, “How to Use Dropout Correctly on Residual Networks with Batch Normalization.” arXiv, Feb. 13, 2023. doi: 10.48550/arXiv.2302.06112.

[95] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, “Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks,” *Sci Rep*, vol. 9, no. 1, Art. no. 1, Nov. 2019, doi: 10.1038/s41598-019-52737-x.

[96] F. H. K. dos S. Tanaka and C. Aranha, “Data Augmentation Using GANs.” arXiv, Apr. 19, 2019. doi: 10.48550/arXiv.1904.09135.

[97] I. Misra and L. van der Maaten, “Self-Supervised Learning of Pretext-Invariant Representations.” arXiv, Dec. 04, 2019. doi: 10.48550/arXiv.1912.01991.

[98] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum Contrast for Unsupervised Visual Representation Learning.” arXiv, Mar. 23, 2020. doi: 10.48550/arXiv.1911.05722.

[99] Centers for Disease Control and Prevention (CDC), “Prevalence of disabilities and associated health conditions among adults--United States, 1999,” *MMWR Morb Mortal Wkly Rep*, vol. 50, no. 7, pp. 120–125, Feb. 2001.

[100] H.-R. Guo *et al.*, “Back pain among workers in the United States: National estimates and workers at high risk,” *American Journal of Industrial Medicine*, vol. 28, no. 5, pp. 591–602, 1995, doi: 10.1002/ajim.4700280504.

[101] H. R. Guo, S. Tanaka, W. E. Halperin, and L. L. Cameron, “Back pain prevalence in US industry and estimates of lost workdays.,” *Am J Public Health*, vol. 89, no. 7, pp. 1029–1035, Jul. 1999, doi: 10.2105/AJPH.89.7.1029.

[102] J. W. Frymoyer and W. L. Cats-Baril, “An Overview of the Incidences and Costs of Low Back Pain,” *Orthopedic Clinics of North America*, vol. 22, no. 2, pp. 263–271, Apr. 1991, doi: 10.1016/S0030-5898(20)31652-7.

[103] J. N. Katz, “Lumbar Disc Disorders and Low-Back Pain: Socioeconomic Factors and Consequences,” *JBJS*, vol. 88, no. suppl_2, pp. 21–24, Apr. 2006, doi: 10.2106/JBJS.E.01273.

[104] T. S. Carey, J. Garrett, A. Jackman, C. McLaughlin, J. Fryer, and D. R. Smucker, “The Outcomes and Costs of Care for Acute Low Back Pain among Patients Seen by Primary Care Practitioners, Chiropractors, and Orthopedic Surgeons,” *New England Journal of Medicine*, vol. 333, no. 14, pp. 913–917, Oct. 1995, doi: 10.1056/NEJM199510053331406.

[105] J. Friedly, L. Chan, and R. Deyo, “Increases in Lumbosacral Injections in the Medicare Population: 1994 to 2001,” *Spine*, vol. 32, no. 16, pp. 1754–1760, Jul. 2007, doi: 10.1097/BRS.0b013e3180b9f96e.

[106] Y. Iizuka *et al.*, “Prevalence of Chronic Nonspecific Low Back Pain and Its Associated Factors among Middle-Aged and Elderly People: An Analysis Based on

Data from a Musculoskeletal Examination in Japan," *Asian Spine J*, vol. 11, no. 6, pp. 989–997, Dec. 2017, doi: 10.4184/asj.2017.11.6.989.

[107] D. I. Rubin, "Epidemiology and Risk Factors for Spine Pain," *Neurologic Clinics*, vol. 25, no. 2, pp. 353–371, May 2007, doi: 10.1016/j.ncl.2007.01.004.

[108] D. Steffens, M. J. Hancock, L. S. M. Pereira, P. M. Kent, J. Latimer, and C. G. Maher, "Do MRI findings identify patients with low back pain or sciatica who respond better to particular interventions? A systematic review," *Eur Spine J*, vol. 25, no. 4, pp. 1170–1187, Apr. 2016, doi: 10.1007/s00586-015-4195-4.

[109] Z. Lidar *et al.*, "Intervertebral Disc Height Changes After Weight Reduction in Morbidly Obese Patients and Its Effect on Quality of Life and Radicular and Low Back Pain," *Spine*, vol. 37, no. 23, pp. 1947–1952, Nov. 2012, doi: 10.1097/BRS.0b013e31825fab16.

[110] A. J. Teichtahl *et al.*, "Modic changes in the lumbar spine and their association with body composition, fat distribution and intervertebral disc height – a 3.0 T-MRI study," *BMC Musculoskelet Disord*, vol. 17, no. 1, p. 92, Feb. 2016, doi: 10.1186/s12891-016-0934-x.

[111] N. Roberts, C. Gratin, and G. H. Whitehouse, "MRI analysis of lumbar intervertebral disc height in young and older populations," *Journal of Magnetic Resonance Imaging*, vol. 7, no. 5, pp. 880–886, 2005, doi: 10.1002/jmri.1880070517.

[112] R. D'hooge, B. Cagnie, G. Crombez, G. Vanderstraeten, M. Dolphens, and L. Danneels, "Increased intramuscular fatty infiltration without differences in lumbar muscle cross-sectional area during remission of unilateral recurrent low back pain,"

Manual Therapy, vol. 17, no. 6, pp. 584–588, Dec. 2012, doi: 10.1016/j.math.2012.06.007.

[113] R. Niemeläinen, M.-M. Briand, and M. C. Battié, “Substantial Asymmetry in Paraspinal Muscle Cross-Sectional Area in Healthy Adults Questions Its Value as a Marker of Low Back Pain and Pathology,” *Spine*, vol. 36, no. 25, pp. 2152–2157, Dec. 2011, doi: 10.1097/BRS.0b013e318204b05a.

[114] S.-Y. Kim, “Changes in Cross-sectional Area of Lumbar Muscle in Patients with Chronic Back Pain,” *The Journal of Korean Physical Therapy*, vol. 22, no. 5, pp. 39–47, 2010.

[115] K. L. Barker, D. R. Shamley, and D. Jackson, “Changes in the Cross-Sectional Area of Multifidus and Psoas in Patients With Unilateral Back Pain: The Relationship to Pain and Disability,” *Spine*, vol. 29, no. 22, p. E515, Nov. 2004, doi: 10.1097/01.brs.0000144405.11661.eb.

[116] F. Rezazadeh, N. Taheri, S. M. Okhravi, and S. M. Hosseini, “The relationship between cross-sectional area of multifidus muscle and disability index in patients with chronic non-specific low back pain,” *Musculoskeletal Science and Practice*, vol. 42, pp. 1–5, Jul. 2019, doi: 10.1016/j.msksp.2019.03.005.

[117] G. Gwak, U. Hwang, S. Jung, H. Kim, J. Kim, and O. Kwon, “Comparison of MRI cross-sectional area and functions of core muscles among asymptomatic individuals with and without lumbar intervertebral disc degeneration,” *BMC Musculoskelet Disord*, vol. 20, no. 1, p. 576, Dec. 2019, doi: 10.1186/s12891-019-2960-y.

[118] H. J. Lee *et al.*, “The Relationship between Cross Sectional Area and Strength of Back Muscles in Patients with Chronic Low Back Pain,” *Ann Rehabil Med*, vol. 36, no. 2, pp. 173–181, Apr. 2012, doi: 10.5535/arm.2012.36.2.173.

[119] A. Suzani, A. Rasoulian, A. Seitel, S. Fels, R. N. Rohling, and P. Abolmaesumi, “Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images,” in *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, International Society for Optics and Photonics, Mar. 2015, p. 941514. doi: 10.1117/12.2081542.

[120] M. Dreischarf, A. Shirazi-Adl, N. Arjmand, A. Rohlmann, and H. Schmidt, “Estimation of loads on human lumbar spine: A review of in vivo and computational model studies,” *Journal of Biomechanics*, vol. 49, no. 6, pp. 833–845, Apr. 2016, doi: 10.1016/j.jbiomech.2015.12.038.

[121] H. Mokhtarzadeh, D. E. Anderson, B. T. Allaire, and M. L. Bouxsein, “Patterns of Load-to-Strength Ratios Along the Spine in a Population-Based Cohort to Evaluate the Contribution of Spinal Loading to Vertebral Fractures,” *Journal of Bone and Mineral Research*, vol. 36, no. 4, pp. 704–711, 2021, doi: 10.1002/jbmr.4222.

[122] A. G. Bruno *et al.*, “Incorporation of CT-based measurements of trunk anatomy into subject-specific musculoskeletal models of the spine influences vertebral loading predictions,” *Journal of Orthopaedic Research*, vol. 35, no. 10, pp. 2164–2173, 2017, doi: 10.1002/jor.23524.

[123] S. L. Delp *et al.*, “OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 11, pp. 1940–1950, Nov. 2007, doi: 10.1109/TBME.2007.901024.

[124] *Sørensen, T. (1948). 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*.

[125] J. Zhou *et al.*, “Automatic Vertebral Body Segmentation Based on Deep Learning of Dixon Images for Bone Marrow Fat Fraction Quantification,” *Frontiers in Endocrinology*, vol. 11, p. 612, 2020, doi: 10.3389/fendo.2020.00612.

[126] W. Xia *et al.*, “Automatic Paraspinal Muscle Segmentation in Patients with Lumbar Pathology Using Deep Convolutional Neural Network,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 318–325. doi: 10.1007/978-3-030-32245-8_36.

[127] C. Iriondo, V. Pedoia, and S. Majumdar, “Lumbar intervertebral disc characterization through quantitative MRI analysis: An automatic voxel-based relaxometry approach,” *Magn Reson Med*, vol. 84, no. 3, pp. 1376–1390, Sep. 2020, doi: 10.1002/mrm.28210.

[128] S. Guinebert, E. Petit, V. Bousson, S. Bodard, N. Amoretti, and B. Kastler, “Automatic semantic segmentation and detection of vertebras and intervertebral discs by neural networks,” *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100055, Jan. 2022, doi: 10.1016/j.cmpbup.2022.100055.

[129] K.-U. Lewandrowskl *et al.*, “Feasibility of Deep Learning Algorithms for Reporting in Routine Spine Magnetic Resonance Imaging,” *Int J Spine Surg*, vol. 14, no. s3, pp. S86–S97, Dec. 2020, doi: 10.14444/7131.

[130] J. Liu *et al.*, “Biomechanical properties of lumbar endplates and their correlation with MRI findings of lumbar degeneration,” *Journal of Biomechanics*, vol. 49, no. 4, pp. 586–593, Feb. 2016, doi: 10.1016/j.jbiomech.2016.01.019.

[131] W. Brinjikji *et al.*, “MRI Findings of Disc Degeneration are More Prevalent in Adults with Low Back Pain than in Asymptomatic Controls: A Systematic Review and Meta-Analysis,” *American Journal of Neuroradiology*, vol. 36, no. 12, pp. 2394–2399, Dec. 2015, doi: 10.3174/ajnr.A4498.

[132] D. Weishaupt *et al.*, “Painful Lumbar Disk Derangement: Relevance of Endplate Abnormalities at MR Imaging,” *Radiology*, vol. 218, no. 2, pp. 420–427, Feb. 2001, doi: 10.1148/radiology.218.2.r01fe15420.

[133] P. Kjaer, L. Korsholm, T. Bendix, J. S. Sorensen, and C. Leboeuf-Yde, “Modic changes and their associations with clinical findings,” *Eur Spine J*, vol. 15, no. 9, pp. 1312–1319, Sep. 2006, doi: 10.1007/s00586-006-0185-x.

[134] M. T. Modic, P. M. Steinberg, J. S. Ross, T. J. Masaryk, and J. R. Carter, “Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging,” *Radiology*, vol. 166, no. 1 Pt 1, pp. 193–199, Jan. 1988, doi: 10.1148/radiology.166.1.3336678.

[135] C. Herlin *et al.*, “Modic changes—Their associations with low back pain and activity limitation: A systematic literature review and meta-analysis,” *PLOS ONE*, vol. 13, no. 8, p. e0200677, Aug. 2018, doi: 10.1371/journal.pone.0200677.

[136] Y.-H. Zhang, C.-Q. Zhao, L.-S. Jiang, X.-D. Chen, and L.-Y. Dai, “Modic changes: a systematic review of the literature,” *Eur Spine J*, vol. 17, no. 10, pp. 1289–1299, Oct. 2008, doi: 10.1007/s00586-008-0758-y.

[137] R. U. Din, X. Cheng, and H. Yang, “Diagnostic Role of Magnetic Resonance Imaging in Low Back Pain Caused by Vertebral Endplate Degeneration,” *Journal of Magnetic Resonance Imaging*, vol. n/a, no. n/a, doi: 10.1002/jmri.27858.

[138] A. J. Fields, E. C. Liebenberg, and J. C. Lotz, “Innervation of pathologies in the lumbar vertebral end plate and intervertebral disc,” *The Spine Journal*, vol. 14, no. 3, pp. 513–521, Mar. 2014, doi: 10.1016/j.spinee.2013.06.075.

[139] S. Ohtori *et al.*, “Tumor Necrosis Factor-Immunoreactive Cells and PGP 9.5-Immunoreactive Nerve Fibers in Vertebral Endplates of Patients With Discogenic Low Back Pain and Modic Type 1 or Type 2 Changes on MRI,” *Spine*, vol. 31, no. 9, pp. 1026–1031, Apr. 2006, doi: 10.1097/01.brs.0000215027.87102.7c.

[140] A. J. Fields *et al.*, “Measuring and reporting of vertebral endplate bone marrow lesions as seen on MRI (Modic changes): recommendations from the ISSLS Degenerative Spinal Phenotypes Group,” *Eur Spine J*, vol. 28, no. 10, pp. 2266–2274, Oct. 2019, doi: 10.1007/s00586-019-06119-6.

[141] T. Bendix, J. S. Sorensen, G. A. C. Henriksson, J. E. Bolstad, E. K. Narvestad, and T. S. Jensen, “Lumbar Modic Changes—A Comparison Between Findings at Low- and High-Field Magnetic Resonance Imaging,” *Spine*, vol. 37, no. 20, pp. 1756–1762, Sep. 2012, doi: 10.1097/BRS.0b013e318257ffce.

[142] A. J. Fields *et al.*, “Measurement of vertebral endplate bone marrow lesion (Modic change) composition with water–fat MRI and relationship to patient-reported outcome measures,” *Eur Spine J*, vol. 30, no. 9, pp. 2549–2556, Sep. 2021, doi: 10.1007/s00586-021-06738-y.

[143] D. C. Karampinos, G. Melkus, T. Baum, J. S. Bauer, E. J. Rummeny, and R. Krug, “Bone marrow fat quantification in the presence of trabecular bone: Initial comparison between water-fat imaging and single-voxel MRS,” *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 1158–1165, 2014, doi: 10.1002/mrm.24775.

[144] A. Biffar, O. Dietrich, S. Sourbron, H.-R. Duerr, M. F. Reiser, and A. Baur-Melnyk, “Diffusion and perfusion imaging of bone marrow,” *European Journal of Radiology*, vol. 76, no. 3, pp. 323–328, Dec. 2010, doi: 10.1016/j.ejrad.2010.03.011.

[145] Y. Wang, T. Videman, R. Niemeläinen, and M. C. Battié, “Quantitative Measures of Modic Changes in Lumbar Spine Magnetic Resonance Imaging: Intra- and Inter-rater Reliability,” *Spine*, vol. 36, no. 15, pp. 1236–1243, Jul. 2011, doi: 10.1097/BRS.0b013e3181ecf283.

[146] N. Lessmann, B. van Ginneken, P. A. de Jong, and I. Išgum, “Iterative fully convolutional neural networks for automatic vertebra segmentation and identification,” *Med Image Anal*, vol. 53, pp. 142–155, Apr. 2019, doi: 10.1016/j.media.2019.02.005.

[147] K. A. Weber *et al.*, “Multi-muscle deep learning segmentation to automate the quantification of muscle fat infiltration in cervical spine conditions,” *Sci Rep*, vol. 11, no. 1, p. 16567, Aug. 2021, doi: 10.1038/s41598-021-95972-x.

[148] Z. Han, B. Wei, A. Mercado, S. Leung, and S. Li, “Spine-GAN: Semantic segmentation of multiple spinal structures,” *Med Image Anal*, vol. 50, pp. 23–35, Dec. 2018, doi: 10.1016/j.media.2018.08.005.

[149] Z. Merali, J. Z. Wang, J. H. Badhiwala, C. D. Witiw, J. R. Wilson, and M. G. Fehlings, “A deep learning model for detection of cervical spinal cord compression in MRI

scans," *Sci Rep*, vol. 11, no. 1, p. 10473, May 2021, doi: 10.1038/s41598-021-89848-3.

[150] J.-T. Lu *et al.*, "DeepSPINE: Automated Lumbar Vertebral Segmentation, Disc-level Designation, and Spinal Stenosis Grading Using Deep Learning," *arXiv:1807.10215 [cs]*, Jul. 2018, Accessed: Oct. 01, 2020. [Online]. Available: http://arxiv.org/abs/1807.10215

[151] J. T. P. D. Hallinan *et al.*, "Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI," *Radiology*, vol. 300, no. 1, pp. 130–138, Jul. 2021, doi: 10.1148/radiol.2021204289.

[152] K. Y. Wang *et al.*, "Using Predictive Modeling and Supervised Machine Learning to Identify Patients at Risk for Venous Thromboembolism Following Posterior Lumbar Fusion," *Global Spine J*, p. 21925682211019360, May 2021, doi: 10.1177/21925682211019361.

[153] P. Fatemi *et al.*, "External validation of a predictive model of adverse events following spine surgery," *Spine J*, pp. S1529-9430(21)00727–0, Jun. 2021, doi: 10.1016/j.spinee.2021.06.006.

[154] A. Jamaludin, T. Kadir, and A. Zisserman, "SpineNet: Automated classification and evidence visualization in spinal MRIs," *Medical Image Analysis*, vol. 41, pp. 63–73, Oct. 2017, doi: 10.1016/j.media.2017.07.002.

[155] H. B. Albert, A. M. Briggs, P. Kent, A. Byrhagen, C. Hansen, and K. Kjaergaard, "The prevalence of MRI-defined spinal pathoanatomies and their association with

Modic changes in individuals seeking care for low back pain," *Eur Spine J*, vol. 20, no. 8, pp. 1355–1362, Aug. 2011, doi: 10.1007/s00586-011-1794-6.

[156] M. Ş. Ekşi *et al.*, "Lumbar intervertebral disc degeneration, end-plates and paraspinal muscle changes in children and adolescents with low-back pain," *J Pediatr Orthop B*, Nov. 2020, doi: 10.1097/BPB.0000000000000833.

[157] D. Mitra, V. N. Cassar-Pullicino, and I. W. McCall, "Longitudinal study of vertebral type-1 end-plate changes on MR of the lumbar spine," *Eur Radiol*, vol. 14, no. 9, pp. 1574–1581, Sep. 2004, doi: 10.1007/s00330-004-2314-4.

[158] M. J. Hutton, J. H. Bayer, and J. M. Powell, "Modic Vertebral Body Changes: The Natural History as Assessed by Consecutive Magnetic Resonance Imaging," *Spine*, vol. 36, no. 26, pp. 2304–2307, Dec. 2011, doi: 10.1097/BRS.0b013e31821604b6.

[159] L. Xu, B. Chu, Y. Feng, F. Xu, and Y.-F. Zou, "Modic changes in lumbar spine: prevalence and distribution patterns of end plate oedema and end plate sclerosis," *Br J Radiol*, vol. 89, no. 1060, p. 20150650, Apr. 2016, doi: 10.1259/bjr.20150650.

[160] V. Andrearczyk and P. F. Whelan, "Using filter banks in Convolutional Neural Networks for texture classification," *Pattern Recognition Letters*, vol. 84, pp. 63–69, Dec. 2016, doi: 10.1016/j.patrec.2016.08.016.

[161] P. Hattikatti, "Texture based interstitial lung disease detection using convolutional neural network," in *2017 International Conference on Big Data, IoT and Data Science (BID)*, Dec. 2017, pp. 18–22. doi: 10.1109/BID.2017.8336567.

[162] T. S. Jensen, J. Karppinen, J. S. Sorensen, J. Niinimäki, and C. Leboeuf-Yde, "Vertebral endplate signal changes (Modic change): a systematic literature review of

prevalence and association with non-specific low back pain," *Eur Spine J*, vol. 17, no. 11, p. 1407, Sep. 2008, doi: 10.1007/s00586-008-0770-2.

[163] H. B. Albert and C. Manniche, "Modic changes following lumbar disc herniation," *Eur Spine J*, vol. 16, no. 7, pp. 977–982, Jul. 2007, doi: 10.1007/s00586-007-0336-8.

[164] P. Kjaer, C. Leboeuf-Yde, L. Korsholm, J. S. Sorensen, and T. Bendix, "Magnetic Resonance Imaging and Low Back Pain in Adults: A Diagnostic Imaging Study of 40-Year-Old Men and Women," *Spine*, vol. 30, no. 10, pp. 1173–1180, May 2005, doi: 10.1097/01.brs.0000162396.97739.76.

[165] M. I. Iosifidis, A. Tsarouhas, and A. Fylaktou, "Lower limb clinical and radiographic osteoarthritis in former elite male athletes," *Knee Surg Sports Traumatol Arthrosc*, vol. 23, no. 9, pp. 2528–2535, Sep. 2015, doi: 10.1007/s00167-014-3047-9.

[166] L. D. Kaplan, M. R. Schurhoff, H. Selesnick, M. Thorpe, and J. W. Uribe, "Magnetic resonance imaging of the knee in asymptomatic professional basketball players," *Arthroscopy*, vol. 21, no. 5, pp. 557–561, May 2005, doi: 10.1016/j.arthro.2005.01.009.

[167] G. P. Pappas, M. A. Vogelsong, E. Staroswiecki, G. E. Gold, and M. R. Safran, "Magnetic Resonance Imaging of Asymptomatic Knees in Collegiate Basketball Players: The Effect of One Season of Play," *Clin J Sport Med*, vol. 26, no. 6, pp. 483–489, Nov. 2016, doi: 10.1097/JSM.0000000000000283.

[168] B. E. Walczak, P. C. McCulloch, R. W. Kang, A. Zelazny, F. Tedeschi, and B. J. Cole, "Abnormal findings on knee magnetic resonance imaging in asymptomatic

NBA players," *J Knee Surg*, vol. 21, no. 1, pp. 27–33, Jan. 2008, doi: 10.1055/s-0030-1247788.

[169] S. J. Matzat, F. Kogan, G. W. Fong, and G. E. Gold, "Imaging Strategies for Assessing Cartilage Composition in Osteoarthritis," *Curr Rheumatol Rep*, vol. 16, no. 11, p. 462, Nov. 2014, doi: 10.1007/s11926-014-0462-3.

[170] R. Crockett, A. Grubelnik, S. Roos, C. Dora, W. Born, and H. Troxler, "Biochemical composition of the superficial layer of articular cartilage," *J Biomed Mater Res A*, vol. 82, no. 4, pp. 958–964, Sep. 2007, doi: 10.1002/jbm.a.31248.

[171] X. Li *et al.*, "IN VIVO T1RHO AND T2 MAPPING OF ARTICULAR CARTILAGE IN OSTEOARTHRITIS OF THE KNEE USING 3 TESLA MRI," *Osteoarthritis Cartilage*, vol. 15, no. 7, pp. 789–797, Jul. 2007, doi: 10.1016/j.joca.2007.01.011.

[172] R. R. Regatte, S. V. S. Akella, J. H. Lonner, J. B. Kneeland, and R. Reddy, "T1rho relaxation mapping in human osteoarthritis (OA) cartilage: comparison of T1rho with T2," *J Magn Reson Imaging*, vol. 23, no. 4, pp. 547–553, Apr. 2006, doi: 10.1002/jmri.20536.

[173] U. Duvvuri, R. Reddy, S. D. Patel, J. H. Kaufman, J. B. Kneeland, and J. S. Leigh, "T1rho-relaxation in articular cartilage: effects of enzymatic degradation," *Magn Reson Med*, vol. 38, no. 6, pp. 863–867, Dec. 1997, doi: 10.1002/mrm.1910380602.

[174] R. R. Regatte, S. V. S. Akella, A. Borthakur, J. B. Kneeland, and R. Reddy, "Proteoglycan depletion-induced changes in transverse relaxation maps of cartilage: comparison of T2 and T1rho," *Acad Radiol*, vol. 9, no. 12, pp. 1388–1394, Dec. 2002, doi: 10.1016/s1076-6332(03)80666-9.

[175] J. van Tiel *et al.*, “Is T1ρ Mapping an Alternative to Delayed Gadolinium-enhanced MR Imaging of Cartilage in the Assessment of Sulphated Glycosaminoglycan Content in Human Osteoarthritic Knees? An in Vivo Validation Study,” *Radiology*, vol. 279, no. 2, pp. 523–531, May 2016, doi: 10.1148/radiol.2015150693.

[176] N. M. Menezes, M. L. Gray, J. R. Hartke, and D. Burstein, “T2 and T1rho MRI in articular cartilage systems,” *Magn Reson Med*, vol. 51, no. 3, pp. 503–509, Mar. 2004, doi: 10.1002/mrm.10710.

[177] U. Duvvuri *et al.*, “Human knee: in vivo T1(rho)-weighted MR imaging at 1.5 T--preliminary experience,” *Radiology*, vol. 220, no. 3, pp. 822–826, Sep. 2001, doi: 10.1148/radiol.2203001662.

[178] R. R. Regatte *et al.*, “3D-T1rho-relaxation mapping of articular cartilage: in vivo assessment of early degenerative changes in symptomatic osteoarthritic subjects,” *Acad Radiol*, vol. 11, no. 7, pp. 741–749, Jul. 2004, doi: 10.1016/j.acra.2004.03.051.

[179] H. F. Atkinson *et al.*, “MRI T2 and T1ρ relaxation in patients at risk for knee osteoarthritis: a systematic review and meta-analysis,” *BMC Musculoskelet Disord*, vol. 20, p. 182, May 2019, doi: 10.1186/s12891-019-2547-7.

[180] X. Li *et al.*, “Simultaneous acquisition of T1ρ and T2 quantification in knee cartilage: repeatability and diurnal variation,” *J Magn Reson Imaging*, vol. 39, no. 5, pp. 1287–1293, May 2014, doi: 10.1002/jmri.24253.

[181] V. Pedoia, X. Li, F. Su, N. Calixto, and S. Majumdar, “Fully automatic analysis of the knee articular cartilage T1ρ relaxation time using voxel-based relaxometry,” *J Magn Reson Imaging*, vol. 43, no. 4, pp. 970–980, Apr. 2016, doi: 10.1002/jmri.25065.

[182] H. Yokota, D. J. Leong, and H. B. Sun, “Mechanical loading: bone remodeling and cartilage maintenance,” *Curr Osteoporos Rep*, vol. 9, no. 4, pp. 237–242, Dec. 2011, doi: 10.1007/s11914-011-0067-y.

[183] R. C. Murray, S. Vedi, H. L. Birch, K. H. Lakhani, and A. E. Goodship, “Subchondral bone thickness, hardness and remodelling are influenced by short-term exercise in a site-specific manner,” *J Orthop Res*, vol. 19, no. 6, pp. 1035–1042, Nov. 2001, doi: 10.1016/S0736-0266(01)00027-4.

[184] Z. Li, S.-Y. Liu, L. Xu, S.-Y. Xu, and G.-X. Ni, “Effects of treadmill running with different intensity on rat subchondral bone,” *Scientific Reports*, vol. 7, May 2017, doi: 10.1038/s41598-017-02126-z.

[185] J. Hashemi *et al.*, “The Geometry of the Tibial Plateau and Its Influence on the Biomechanics of the Tibiofemoral Joint,” *J Bone Joint Surg Am*, vol. 90, no. 12, pp. 2724–2734, Dec. 2008, doi: 10.2106/JBJS.G.01358.

[186] S. J. Shultz and R. J. Schmitz, “Tibial plateau geometry influences lower extremity biomechanics during landing,” *Am J Sports Med*, vol. 40, no. 9, pp. 2029–2036, Sep. 2012, doi: 10.1177/0363546512453295.

[187] H. C. Smith *et al.*, “Risk factors for anterior cruciate ligament injury: a review of the literature - part 1: neuromuscular and anatomic risk,” *Sports Health*, vol. 4, no. 1, pp. 69–78, Jan. 2012, doi: 10.1177/1941738111428281.

[188] B. Dube, M. A. Bowes, E. M. A. Hensor, A. Barr, S. R. Kingsbury, and P. G. Conaghan, “The relationship between two different measures of osteoarthritis bone pathology, bone marrow lesions and 3D bone shape: data from the Osteoarthritis

Initiative," *Osteoarthritis Cartilage*, vol. 26, no. 10, pp. 1333–1337, Oct. 2018, doi: 10.1016/j.joca.2018.06.011.

[189] V. Pedoia *et al.*, "Three-dimensional MRI-based statistical shape model and application to a cohort of knees with acute ACL injury," *Osteoarthritis and Cartilage*, vol. 23, no. 10, pp. 1695–1703, Oct. 2015, doi: 10.1016/j.joca.2015.05.027.

[190] J. S. Gregory, R. J. Barr, K. Yoshida, S. Alesci, D. M. Reid, and R. M. Aspden, "Statistical shape modelling provides a responsive measure of morphological change in knee osteoarthritis over 12 months," *Rheumatology*, vol. 59, no. 9, pp. 2419–2426, Sep. 2020, doi: 10.1093/rheumatology/kez610.

[191] D. J. Haverkamp, D. Schiphof, S. M. Bierma-Zeinstra, H. Weinans, and J. H. Waarsing, "Variation in joint shape of osteoarthritic knees," *Arthritis Rheum*, vol. 63, no. 11, pp. 3401–3407, Nov. 2011, doi: 10.1002/art.30575.

[192] T. L. Bredbenner, T. D. Eliason, R. S. Potter, R. L. Mason, L. M. Havill, and D. P. Nicolella, "Statistical shape modeling describes variation in tibia and femur surface geometry between Control and Incidence groups from the osteoarthritis initiative database," *J Biomech*, vol. 43, no. 9, pp. 1780–1786, Jun. 2010, doi: 10.1016/j.jbiomech.2010.02.015.

[193] K. Young *et al.*, "Advanced Knee Imaging Study in NCAA Division 1 Basketball: Protocol Development and Preliminary Results," presented at the ISMRM 2018, Jun. 2018.

[194] T. Hartkens, D. Rueckert, J. A. Schnabel, D. J. Hawkes, and D. L. G. Hill, "VTK CISG Registration Toolkit An Open Source Software Package for Affine and Non-rigid Registration of Single- and Multimodal 3D Images," in *Bildverarbeitung für die*

Medizin 2002, M. Meiler, D. Saupe, F. Kruggel, H. Handels, and T. M. Lehmann, Eds., in Informatik aktuell. Berlin, Heidelberg: Springer, 2002, pp. 409–412. doi: 10.1007/978-3-642-55983-9_96.

[195] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, “elastix: a toolbox for intensity-based medical image registration,” *IEEE Trans Med Imaging*, vol. 29, no. 1, pp. 196–205, Jan. 2010, doi: 10.1109/TMI.2009.2035616.

[196] J. Carballido-Gamio *et al.*, “Inter-subject comparison of MRI knee cartilage thickness,” *Med Image Anal*, vol. 12, no. 2, pp. 120–135, Apr. 2008, doi: 10.1016/j.media.2007.08.002.

[197] W. Lorensen and H. Cline, “Marching Cubes: A High Resolution 3D Surface Construction Algorithm,” *ACM SIGGRAPH Computer Graphics*, vol. 21, p. 163, Aug. 1987, doi: 10.1145/37401.37422.

[198] H. Lombaert, L. Grady, J. R. Polimeni, and F. Cheriet, “FOCUSR: Feature Oriented Correspondence Using Spectral Regularization–A Method for Precise Surface Matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2143–2160, Sep. 2013, doi: 10.1109/TPAMI.2012.276.

[199] C. S. Paranjape *et al.*, “A New Stress Test for Knee Joint Cartilage,” *Sci Rep*, vol. 9, no. 1, Art. no. 1, Feb. 2019, doi: 10.1038/s41598-018-38104-2.

[200] B. L. Wise *et al.*, “Patterns of compartment involvement in tibiofemoral osteoarthritis in men and women and in whites and African Americans,” *Arthritis Care Res (Hoboken)*, vol. 64, no. 6, pp. 847–852, Jun. 2012, doi: 10.1002/acr.21606.

[201] K. R. Vincent, B. P. Conrad, B. J. Fregly, and H. K. Vincent, “The pathophysiology of osteoarthritis: a mechanical perspective on the knee joint,” *PM R*, vol. 4, no. 5 Suppl, pp. S3-9, May 2012, doi: 10.1016/j.pmrj.2012.01.020.

[202] A. J. Sophia Fox, A. Bedi, and S. A. Rodeo, “The basic science of articular cartilage: structure, composition, and function,” *Sports Health*, vol. 1, no. 6, pp. 461–468, Nov. 2009, doi: 10.1177/1941738109350438.

[203] A. M. J. Bull, P. H. Earnshaw, A. Smith, M. V. Katchburian, A. N. A. Hassan, and A. A. Amis, “Intraoperative measurement of knee kinematics in reconstruction of the anterior cruciate ligament,” *J Bone Joint Surg Br*, vol. 84, no. 7, pp. 1075–1081, Sep. 2002, doi: 10.1302/0301-620x.84b7.13094.

[204] A. E. Wluka, Y. Wang, S. R. Davis, and F. M. Cicuttini, “Tibial plateau size is related to grade of joint space narrowing and osteophytes in healthy women and in women with osteoarthritis,” *Ann Rheum Dis*, vol. 64, no. 7, pp. 1033–1037, Jul. 2005, doi: 10.1136/ard.2004.029082.

[205] S. Donnelly, D. J. Hart, D. V. Doyle, and T. D. Spector, “Spiking of the tibial tubercles--a radiological feature of osteoarthritis?,” *Ann Rheum Dis*, vol. 55, no. 2, pp. 105–108, Feb. 1996.

[206] M. R. Hayeri, M. Shiehmorteza, D. J. Trudell, T. Hefflin, and D. Resnick, “Proximal tibial osteophytes and their relationship with the height of the tibial spines of the intercondylar eminence: paleopathological study,” *Skeletal Radiol*, vol. 39, no. 9, pp. 877–881, 2010, doi: 10.1007/s00256-009-0838-z.

[207] G. Wibeeg, “Roentgenographs and Anatomic Studies on the Femoropatellar Joint: With Special Reference to Chondromalacia Patellae,” *Acta Orthopaedica*

Scandinavica, vol. 12, no. 1–4, pp. 319–410, Jan. 1941, doi: 10.3109/17453674108988818.

[208] N. S. C. and O. Branch, "Osteoarthritis Initiative," *National Institute of Arthritis and Musculoskeletal and Skin Diseases*, Mar. 08, 2017. https://www.niams.nih.gov/grants-funding/funded-research/osteoarthritis-initiative (accessed Apr. 29, 2023).

[209] C. Iriondo, F. Liu, F. Calivà, S. Kamat, S. Majumdar, and V. Pedoia, "Towards understanding mechanistic subgroups of osteoarthritis: 8-year cartilage thickness trajectory analysis," *Journal of Orthopaedic Research*, vol. 39, no. 6, pp. 1305–1317, 2021, doi: 10.1002/jor.24849.

[210] A. Morales Martinez *et al.*, "Learning osteoarthritis imaging biomarkers from bone surface spherical encoding," *Magnetic Resonance in Medicine*, vol. 84, no. 4, pp. 2190–2203, 2020, doi: 10.1002/mrm.28251.

[211] F. Calivá, S. Kamat, A. Morales Martinez, S. Majumdar, and V. Pedoia, "Surface spherical encoding and contrastive learning for virtual bone shape aging," *Med Image Anal*, vol. 77, p. 102388, Apr. 2022, doi: 10.1016/j.media.2022.102388.

[212] M. Cross *et al.*, "The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study," *Annals of the Rheumatic Diseases*, vol. 73, no. 7, pp. 1323–1330, Jul. 2014, doi: 10.1136/annrheumdis-2013-204763.

[213] F. W. Roemer *et al.*, "MRI-based screening for structural definition of eligibility in clinical DMOAD trials: Rapid OsteoArthritis MRI Eligibility Score (ROAMES)," *Osteoarthritis and Cartilage*, vol. 28, no. 1, pp. 71–81, Jan. 2020, doi: 10.1016/j.joca.2019.08.005.

[214] N. G. Shrive, J. J. O'Connor, and J. W. Goodfellow, "Load-bearing in the knee joint," *Clin Orthop Relat Res*, no. 131, pp. 279–287, Apr. 1978.

[215] H. Kurosawa, T. Fukubayashi, and H. Nakajima, "Load-bearing mode of the knee joint: physical behavior of the knee joint with or without menisci," *Clin Orthop Relat Res*, no. 149, pp. 283–290, Jun. 1980.

[216] T. M. Link *et al.*, "Osteoarthritis: MR Imaging Findings in Different Stages of Disease and Correlation with Clinical Findings," *Radiology*, vol. 226, no. 2, pp. 373–381, Feb. 2003, doi: 10.1148/radiol.2262012190.

[217] M. Englund, F. W. Roemer, D. Hayashi, M. D. Crema, and A. Guermazi, "Meniscus pathology, osteoarthritis and the treatment controversy," *Nat Rev Rheumatol*, vol. 8, no. 7, Art. no. 7, Jul. 2012, doi: 10.1038/nrrheum.2012.69.

[218] M.-J. Berthiaume *et al.*, "Meniscal tear and extrusion are strongly associated with progression of symptomatic knee osteoarthritis as assessed by quantitative magnetic resonance imaging," *Annals of the Rheumatic Diseases*, vol. 64, no. 4, pp. 556–563, Apr. 2005, doi: 10.1136/ard.2004.023796.

[219] D. R. Gale, C. E. Chaisson, S. M. S. Totterman, R. K. Schwartz, M. E. Gale, and D. Felson, "Meniscal subluxation: association with osteoarthritis and joint space narrowing," *Osteoarthritis and Cartilage*, vol. 7, no. 6, pp. 526–532, Nov. 1999, doi: 10.1053/joca.1999.0256.

[220] M. Englund *et al.*, "Meniscal tear in knees without surgery and the development of radiographic osteoarthritis among middle-aged and elderly persons: The Multicenter Osteoarthritis Study," *Arthritis Rheum*, vol. 60, no. 3, pp. 831–839, Mar. 2009, doi: 10.1002/art.24383.

[221] L. Torres *et al.*, “The relationship between specific tissue lesions and pain severity in persons with knee osteoarthritis,” *Osteoarthritis and Cartilage*, vol. 14, no. 10, pp. 1033–1040, Oct. 2006, doi: 10.1016/j.joca.2006.03.015.

[222] P. R. Kornaat *et al.*, “MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system,” *Skeletal Radiol*, vol. 34, no. 2, pp. 95–102, Feb. 2005, doi: 10.1007/s00256-004-0828-0.

[223] W. Wirth *et al.*, “A three-dimensional quantitative method to measure meniscus shape, position, and signal intensity using MR images: A pilot study and preliminary results in knee osteoarthritis,” *Magnetic Resonance in Medicine*, vol. 63, no. 5, pp. 1162–1171, 2010, doi: 10.1002/mrm.22380.

[224] T. Kawahara, T. Sasho, J. Katsuragi, T. Ohnishi, and H. Haneishi, “Relationship between knee osteoarthritis and meniscal shape in observation of Japanese patients by using magnetic resonance imaging,” *Journal of Orthopaedic Surgery and Research*, vol. 12, no. 1, p. 97, Jun. 2017, doi: 10.1186/s13018-017-0595-y.

[225] A. Wenger *et al.*, “Meniscus Body Position, Size, and Shape in Persons With and Persons Without Radiographic Knee Osteoarthritis: Quantitative Analyses of Knee Magnetic Resonance Images From the Osteoarthritis Initiative,” *Arthritis & Rheumatism*, vol. 65, no. 7, pp. 1804–1811, 2013, doi: 10.1002/art.37947.

[226] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, “Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative,” *Medical Image Analysis*, vol. 52, pp. 109–118, Feb. 2019, doi: 10.1016/j.media.2018.11.009.

[227] M. Byra *et al.*, “Knee menisci segmentation and relaxometry of 3D ultrashort echo time cones MR imaging using attention U-Net with transfer learning,” *Magnetic Resonance in Medicine*, vol. 83, no. 3, pp. 1109–1122, 2020, doi: 10.1002/mrm.27969.

[228] A. D. Desai *et al.*, “The International Workshop on Osteoarthritis Imaging Knee MRI Segmentation Challenge: A Multi-Institute Evaluation and Analysis Framework on a Standardized Dataset,” *Radiol Artif Intell*, vol. 3, no. 3, p. e200078, Feb. 2021, doi: 10.1148/ryai.2021200078.

[229] S. Gaj, M. Yang, K. Nakamura, and X. Li, “Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks,” *Magnetic Resonance in Medicine*, vol. 84, no. 1, pp. 437–449, 2020, doi: 10.1002/mrm.28111.

[230] B. Norman, V. Pedoia, and S. Majumdar, “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*, vol. 288, no. 1, pp. 177–185, Jul. 2018, doi: 10.1148/radiol.2018172322.

[231] E. Panfilov, A. Tiulpin, M. T. Nieminen, S. Saarakkala, and V. Casula, “Deep learning-based segmentation of knee MRI for fully automatic subregional morphological assessment of cartilage tissues: Data from the Osteoarthritis Initiative,” *Journal of Orthopaedic Research*, vol. 40, no. 5, pp. 1113–1124, 2022, doi: 10.1002/jor.25150.

[232] Z. Zhou, G. Zhao, R. Kijowski, and F. Liu, “Deep convolutional neural network for segmentation of knee joint anatomy,” *Magnetic Resonance in Medicine*, vol. 80, no. 6, pp. 2759–2770, 2018, doi: 10.1002/mrm.27229.

[233] F. Ambellan, H. Lamecker, C. von Tycowicz, and S. Zachow, “Statistical Shape Models: Understanding and Mastering Variation in Anatomy,” in *Biomedical Visualisation : Volume 3*, P. M. Rea, Ed., in Advances in Experimental Medicine and Biology. Cham: Springer International Publishing, 2019, pp. 67–84. doi: 10.1007/978-3-030-19385-0_5.

[234] K. T. Gao *et al.*, “Multiparametric MRI characterization of knee articular cartilage and subchondral bone shape in collegiate basketball players,” *Journal of Orthopaedic Research*, vol. 39, no. 7, pp. 1512–1522, 2021, doi: 10.1002/jor.24851.

[235] B. L. Wise *et al.*, “Bone shape mediates the relationship between sex and incident knee osteoarthritis,” *BMC Musculoskeletal Disorders*, vol. 19, no. 1, p. 331, Sep. 2018, doi: 10.1186/s12891-018-2251-z.

[236] M. A. Bowes *et al.*, “Machine-learning, MRI bone shape and important clinical outcomes in osteoarthritis: data from the Osteoarthritis Initiative,” *Annals of the Rheumatic Diseases*, vol. 80, no. 4, pp. 502–508, Apr. 2021, doi: 10.1136/annrheumdis-2020-217160.

[237] A. C. T. Vrancken *et al.*, “3D geometry analysis of the medial meniscus – a statistical shape modeling approach,” *Journal of Anatomy*, vol. 225, no. 4, pp. 395–402, 2014, doi: 10.1111/joa.12223.

[238] B. Dube, M. A. Bowes, S. R. Kingsbury, E. M. A. Hensor, S. Muzumdar, and P. G. Conaghan, “Where does meniscal damage progress most rapidly? An analysis using

three-dimensional shape models on data from the Osteoarthritis Initiative," *Osteoarthritis and Cartilage*, vol. 26, no. 1, pp. 62–71, Jan. 2018, doi: 10.1016/j.joca.2017.10.012.

[239] F. Eckstein, W. Wirth, and M. C. Nevitt, "Recent advances in osteoarthritis imaging—the Osteoarthritis Initiative," *Nat Rev Rheumatol*, vol. 8, no. 10, Art. no. 10, Oct. 2012, doi: 10.1038/nrrheum.2012.113.

[240] C. G. Peterfy, E. Schneider, and M. Nevitt, "The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee," *Osteoarthritis and Cartilage*, vol. 16, no. 12, pp. 1433–1441, Dec. 2008, doi: 10.1016/j.joca.2008.06.016.

[241] G. S. Van Thiel, N. Verma, A. Yanke, S. Basu, J. Farr, and B. Cole, "Meniscal Allograft Size Can Be Predicted by Height, Weight, and Gender," *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, vol. 25, no. 7, pp. 722–727, Jul. 2009, doi: 10.1016/j.arthro.2009.01.004.

[242] K. Bloecker *et al.*, "Revision 1 Size and position of the healthy meniscus, and its Correlation with sex, height, weight, and bone area- a cross-sectional study," *BMC Musculoskeletal Disorders*, vol. 12, no. 1, p. 248, Oct. 2011, doi: 10.1186/1471-2474-12-248.

[243] O. C. Brantigan and A. F. Voshell, "THE MECHANICS OF THE LIGAMENTS AND MENISCI OF THE KNEE JOINT," *JBJS*, vol. 23, no. 1, pp. 44–66, Jan. 1941.

[244] W. O. Thompson, F. L. Thaete, F. H. Fu, and S. F. Dye, "Tibial meniscal dynamics using three-dimensional reconstruction of magnetic resonance images," *Am J Sports Med*, vol. 19, no. 3, pp. 210–216, May 1991, doi: 10.1177/036354659101900302.

[245] M. Yaniv and N. Blumberg, “The discoid meniscus,” *J Child Orthop*, vol. 1, no. 2, pp. 89–96, Jul. 2007, doi: 10.1007/s11832-007-0029-1.

[246] X. Mei *et al.*, “RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning,” *Radiol Artif Intell*, vol. 4, no. 5, p. e210315, Sep. 2022, doi: 10.1148/ryai.210315.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

627DEF7AA0E64F0... Author Signature

5/25/2023
Date