University of Antwerp

3D knee morphology analysis and predictive models for data-driven patient selection: shaping the future

Development of a prognostic model for arthroscopic partial medial meniscectomy

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Medical Sciences and Doctor of Science: Physics
at the University of Antwerp

Jonas Grammens

Supervisors
dr. F. Danckaers
dr. A. Van Haver
prof. dr. J. Sijbers
prof. dr. P. Verdonk

Antwerpen, 2024

Jury

Chair
prof. dr. A. Snoeckx, University of Antwerp, Belgium

Supervisors
dr. F. Danckaers, University of Antwerp, Belgium
dr. A. Van Haver, University of Antwerp, Belgium
prof. dr. J. Sijbers, University of Antwerp, Belgium
prof. dr. P. Verdonk, University of Antwerp, Belgium

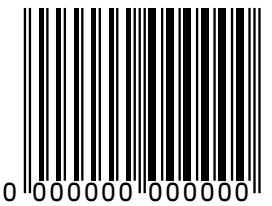
Members
prof. dr. F Van Glabbeek , University of Antwerp, Belgium
prof. dr. L. Scheys, Katholieke Universiteit Leuven, Belgium
prof. dr. J. Vandemeulebroucke, Vrije Universiteit Brussel, Belgium

Contact
Jonas Grammens

University of Antwerp
Faculty of Medicine, Antwerp Surgical Training, Anatomy and Research Centre
Faculty of Science, imec-VisionLab
Universiteitsplein 1, 2650 Wilrijk, België
jonas.grammens@uantwerpen.be

© 2024 Jonas Grammens
All rights reserved.

ISBN 000-00-00000-00-0
Wettelijk depot D/2024/00.000/00



Dutch title:

3D vormanalyse van de knie en predictieve modellen voor data-ondersteunde patiëntselectie

Ontwikkeling van een prognostisch model voor arthroscopische partiële mediale meniscectomie

"All models are wrong,
but some of them are useful.

GEORGE E.P. BOX, British statistician, 1976

To my parents, in loving memory.

Acknowledgments

This PhD journey has been a hell of a ride. First and foremost, I want to thank my promotors, family, and friends for their understanding, advice, and unwavering support, especially during the challenging times in my personal life. Thankfully, during this journey I also encountered moments of pure joy, such as our wedding and the birth of our son, Marcel. Thank you, Marcel, for brightening our days with your newfound ability to chat ("Papa nie werken bureau", accompanied by your charming little hand gestures).

This research project is truly the result of teamwork. I am therefore deeply grateful to Femke Danckaers and Annemieke Van Haver, for guiding me through the wonderful world of Linux, C++ and Python programming, statistical shape analysis, appropriate validation methods, and academic writing. To my promotors Peter Verdonk and Jan Sijbers, thank you for your unlimited and contagious enthusiasm, critical attitude and never ending attempts to teach me the noble art of scientific storytelling. I am also thankful to the MEFISTO consortium partners, particularly Girish, Lotem, Elizaveta, and Peter, for their invaluable input and support during the data collection for this project.

I extend my gratitude to my Visionlab colleagues, Brian, Pavel, and Julian, for creating a supportive and inspiring work environment, despite the shift away from regular campus commutes even some years after the COVID pandemic. My thanks also go to Shabab and especially Luis for their most useful guidance in automated image segmentation using deep learning techniques. I would also like to thank my fellow researchers from Ghent; the EORS conferences were always a highlight, providing both scientific and cultural enrichment.

Finally, I want to extend heartfelt thanks to my close family (in-law) members—Marijke, Christel, Peter, Nicholas, Gilles, Maaike, Lieven, Femke, Dempsey, Bernd, and Jozefien—and to my friends, for their support and understanding. Whether it was tolerating me bringing a computer along on weekend trips to wrap up a manuscript or experiment, or stepping in for last-minute babysitting with Marcel, your help and support are greatly appreciated. Last but not least, my deepest gratitude goes to my wife, Aline, for her patience and understanding, especially as my work-life balance became almost nonexistent towards the end of this journey.

Contents

Summary xv

Samenvatting xvii

I Clinical background and research objectives 1

1 Anatomy of the knee 3

1.1 Anatomical reference frame . . . 3

1.2 Bony anatomy of the knee . . . 4

1.2.1 Articular cartilage . . . 8

1.3 Soft-tissue enveloppe . . . 8

1.3.1 Knee capsule . . . 8

1.3.2 Muscles . . . 9

1.3.3 Ligaments . . . 10

1.3.4 Menisci . . . 11

1.4 Elementary knee biomechanics . . . 12

1.4.1 Kinematics of the knee . . . 12

1.4.2 Coronal alignment of the lower limb . . . 13

Bibliography . . . 15

2 Degenerative meniscus and knee osteoarthritis 19

2.1 Introduction . . . 19

2.2 Degenerative meniscus lesions . . . 20

2.2.1 Epidemiology and pathophysiology . . . 20

2.2.2 Treatment options for degenerative meniscus lesions 21
2.3 Post-meniscectomy syndrome 22
2.3.1 Epidemiology and pathophysiology 22
2.3.2 Treatment options for post-meniscectomy syndrome 23
2.3.3 Prevention of the post-meniscectomy syndrome 25
2.4 Knee osteoarthritis in the meniscus-deficient knee 25
2.4.1 Epidemiology and pathophysiology 25
2.4.2 Delay of post-meniscectomy knee osteoarthritis 25
2.4.3 Treatment options for post-meniscectomy knee osteoarthritis . . . 26
2.5 Consolidating mechanical risk factors in knee OA 27
Bibliography 28

3 Research objectives **33**
3.1 Clinical objectives 33
3.1.1 3D knee morphology as risk factor for symptomatic knee degeneration in the meniscus deficient knee 33
3.1.2 Optimal patient selection for arthroscopic partial medial meniscectomy 34
3.2 Technological framework 35
3.2.1 Automated image segmentation 35
3.2.2 Automated anatomical landmarking 35
3.2.3 Automated statistical shape analysis 35
3.2.4 Automated prediction of treatment response versus failure 36
3.3 Outline of the thesis 36

II Methodological framework **37**

4 Image modalities **39**
4.1 Computed Tomography 39
4.1.1 Basic principles of CT imaging 39

4.1.2 Challenges in CT imaging . . . 40
4.1.3 Applications of CT in orthopaedics . . . 41
4.1.4 Conclusion . . . 41
4.2 MRI . . . 42
4.2.1 Basic principles of MRI . . . 42
4.2.2 Thick-Slice 2D MRI / Thin-Slice 3D MRI . . . 44
4.2.3 Challenges in MRI Imaging . . . 45
4.2.4 Applications of MRI in Orthopaedics . . . 46
4.2.5 Conclusion . . . 47
4.3 Image modalities summary . . . 47
Bibliography . . . 48

5 Image segmentation and 3D meshing 49
5.1 Manual image segmentation . . . 49
5.1.1 Introduction . . . 49
5.1.2 Software for manual image segmentation . . . 50
5.1.3 Challenges and pitfalls . . . 51
5.2 3D meshing . . . 53
Bibliography . . . 55

6 Automated image segmentation: neural networks 57
6.1 Introduction . . . 57
6.2 Neural networks . . . 58
6.2.1 Deep learning . . . 58
6.3 U-Net . . . 59
6.4 nnU-Net . . . 61
6.4.1 Automated configuration . . . 61
6.4.2 Predefined network architectures: 2D and 3D approaches . . . 62
6.4.3 Training and optimization . . . 62
6.4.4 Conclusion . . . 62

6.5 Swin-UNetR . . . 64
6.5.1 Transformers . . . 64
6.5.2 Swin Transformer . . . 64
6.5.3 Swin-UnetR: A Transformer-Based Medical Image Segmentation Model . . . 64
6.5.4 Conclusion . . . 66
6.6 MedNeXt . . . 66
6.6.1 From ConvNeXt... . . . 66
6.6.2 ...to MedNeXt . . . 67
6.7 SegMamba . . . 67
6.7.1 Introduction . . . 67
6.7.2 State space models . . . 68
6.7.3 Key architecture features of SegMamba . . . 68
6.7.4 Conclusion . . . 70
Bibliography . . . 70

7 Automated Landmark Annotation for Morphometric Analysis of Distal Femur and Proximal Tibia 73

7.1 Introduction . . . 73
7.2 Methods . . . 74
7.2.1 Data and workflow . . . 74
7.2.2 Manual landmark annotation . . . 75
7.2.3 Automated landmark annotation . . . 75
7.2.4 Observations . . . 76
7.2.5 Validation study for manual morphometric analysis . . . 80
7.2.6 Validation study for automated morphometric analysis . . . 81
7.2.7 Time consumption . . . 81
7.2.8 Statistical analysis . . . 81
7.3 Results . . . 82
7.3.1 Validation study for manual morphometric analysis . . . 82

7.3.2 Validation study for automated morphometric analysis 84
7.3.3 Time consumption 90
7.4 Discussion 90
7.5 Conclusion 92
Bibliography 93

8 Statistical shape analysis: technical background 95
8.1 Introduction 95
8.2 Dataset preparation 95
8.2.1 Procrustes analysis 96
8.3 Surface registration 97
8.3.1 Initial alignment 97
8.3.2 Corresponding point search 98
8.3.3 Rigid transformation 98
8.3.4 Elasticity modulated registration 99
8.3.5 Shape model prior modulated registration 99
8.4 Constructing a statistical shape model (SSM) 100
8.4.1 Principal component analysis (PCA) 101
8.4.2 PPSA 103
8.5 Model performance assessment 104
8.6 Deep learning: the future? 105
Bibliography 106

9 Elementary concepts of machine learning 109
9.1 Introduction 109
9.2 Dataset preprocessing 111
9.2.1 Manual data checks and the role of domain expertise 111
9.2.2 Data cleaning 112
9.2.3 Feature scaling 112
9.2.4 Feature encoding 112

9.2.5 Feature selection . . . 112
9.2.6 Data transformation and augmentation . . . 113
9.3 Data splits: train, validation and test sets . . . 113
9.4 Hyperparameter optimization . . . 114
9.5 Model training . . . 115
9.5.1 Avoiding overfitting . . . 115
9.6 Model evaluation . . . 115
9.6.1 Evaluation Metrics . . . 116
9.6.2 Evaluation Strategies . . . 117
9.7 Conclusion . . . 118
Bibliography . . . 118

III Clinical applications 120

10 Statistical parametric mapping for segmentation evaluation 121
10.1 Introduction . . . 121
10.2 Materials and methods . . . 123
10.2.1 Dataset description . . . 123
10.2.2 Image preprocessing . . . 123
10.2.3 Model training . . . 124
10.2.4 Model inference . . . 125
10.2.5 Conventional segmentation accuracy metrics . . . 125
10.2.6 3D post-processing . . . 126
10.2.7 Surface registration . . . 126
10.2.8 Descriptive statistics of segmentation errors . . . 126
10.3 Results . . . 127
10.3.1 Conventional segmentation accuracy metrics . . . 127
10.3.2 3D surface mapping . . . 127
10.3.3 Hardware requirements and inference time . . . 135

10.4 Discussion . . . 136
10.5 Conclusion . . . 139
Bibliography . . . 140

11 Automated anatomical landmark-based applications 143

11.1 Application: Femoral anteversion measurement . . . 143
11.1.1 Introduction . . . 143
11.1.2 Materials and methods . . . 144
11.1.3 Results . . . 146
11.1.4 Discussion . . . 146
11.2 Application: Meniscal Anatomy Estimation . . . 148
11.2.1 Introduction . . . 148
11.2.2 Materials and methods . . . 148
11.2.3 Results . . . 150
11.2.4 Discussion . . . 151
Bibliography . . . 152

12 Small medial femoral condyle: a pilot study 155

12.1 Introduction . . . 155
12.2 Materials and methods . . . 156
12.2.1 Study population . . . 158
12.2.2 Generation and isotropic scaling of 3D computer models . . . 158
12.2.3 Definition of the landmarks . . . 158
12.2.4 Definition of the reference planes . . . 159
12.2.5 Measurements . . . 159
12.2.6 Statistical analysis . . . 159
12.2.7 Global shape analysis . . . 160
12.3 Results . . . 160
12.3.1 Medial compartment degeneration . . . 160

12.3.2 Landmark-based analysis . . . 161
12.3.3 Global shape analysis . . . 163
12.4 Discussion . . . 163
12.5 Conclusion . . . 168
Bibliography . . . 168

13 Meniscus Size Differs Between Patient and Donor Populations for Meniscus Allograft Transplantation 173

13.1 Introduction . . . 173
13.2 Materials and methods . . . 174
13.2.1 Data collection and analysis . . . 174
13.2.2 Distribution of meniscus size and area . . . 175
13.2.3 Body Mass to Meniscus Index (BMMI) and Height Over Meniscus Size Index (HMI) . . . 175
13.2.4 Time to Matching Evaluation . . . 175
13.2.5 Statistical analysis . . . 176
13.3 Results . . . 176
13.3.1 Medial meniscus (donor versus patient populations) . . . 177
13.3.2 Lateral meniscus (donor versus patient populations) . . . 179
13.3.3 Time to Match a Patient Request to Donor Graft . . . 181
13.4 Discussion . . . 181
13.5 Conclusion . . . 184
Bibliography . . . 185

14 3D bone morphology is a risk factor for medial post-meniscectomy syndrome: a retrospective cohort study 187

14.1 Introduction . . . 187
14.2 Materials and methods . . . 188
14.2.1 Patient selection and study design . . . 189
14.2.2 Data collection . . . 190
14.2.3 Generation of patient-specific 3D bone models . . . 190

14.2.4 Statistical shape model: data-driven morphology description . . . 191
14.2.5 Statistical analysis . . . 193
14.3 Results . . . 194
14.3.1 Study population . . . 194
14.3.2 Knee morphology comparison between R and MPMS group . . . 196
14.3.3 Correlation analysis between demographic, clinical and knee morphology variables . . . 199
14.3.4 Prediction of response to treatment based on knee morphology . . 200
14.4 Discussion . . . 201
14.5 Conclusion . . . 203
Bibliography . . . 203

15 Machine learning for APMM outcome prediction: clinical and morphological data from the MEFISTO project 209

15.1 Introduction . . . 209
15.2 Materials and methods . . . 211
15.2.1 Study design and data collection . . . 211
15.2.2 Knee shape characterization: surface registration for statistical shape analysis . . . 213
15.2.3 Predictive model . . . 213
15.2.4 Model explainability . . . 216
15.3 Results . . . 217
15.3.1 Clinical features . . . 217
15.3.2 Predictive model . . . 217
15.3.3 Model explainability . . . 220
15.4 Discussion . . . 223
15.5 Conclusion . . . 225
Bibliography . . . 225

IV Conclusion and future perspectives 230

16 Conclusions 231

16.1 Technological findings . . . 231

16.1.1 Automated image segmentation . . . 231

16.1.2 Automated anatomical landmarking . . . 231

16.1.3 Automated statistical shape analysis . . . 232

16.1.4 Automated prediction of response to treatment versus failure . . . 232

16.2 Clinical findings . . . 232

16.2.1 Clinical applications of 3D image analysis . . . 232

16.2.2 3D knee morphology as risk factor for symptomatic knee degeneration in the meniscus deficient knee . . . 233

16.2.3 Optimal patient selection for arthroscopic partial medial meniscectomy . . . 233

16.3 Closing remark . . . 234

17 Future perspectives 235

17.1 Technological perspectives . . . 235

17.2 Clinical perspectives . . . 236

Bibliography . . . 237

List of Figures I

List of Tables V

A Appendix IX

B Curriculum vitae XI

C List of Publications XIII

C.1 Articles . . . XIII

C.2 Conferences . . . XIV

Summary

Multiple risk factors contributing to the onset and progression of knee osteoarthritis (OA) have already been described in current literature, including the loss of a functional meniscus. However, not all cases of early onset knee OA can be explained by the currently known risk factors. Originating from an empirical observation in clinics, the bony knee morphology is hypothesized to play a role in this multifactorial degenerative condition. Innovations in biomedical imaging and AI computer vision algorithms support the automation and scalability of 3D anatomical modeling. Next, automated landmarking methods enable precise and transparent morphometric analysis of the knee. Unique morphological traits, such as a narrow medial femoral condyle, appear linked to medial knee degeneration, while size discrepancies in meniscal allograft transplantation between donor supply (larger medial meniscal size) and acceptor demand (smaller medial meniscal size) highlight challenges in matching donors with recipients. Statistical shape analysis allows to compare the main modes of shape variation between responders to arthroscopic partial medial meniscectomy (APMM) and patients with the medial post-meniscectomy syndrome (MPMS; clinical diagnosis, characterized by recurrent pain related to the meniscal deficiency).

In this PhD thesis, a prognostic model was developed for arthroscopic partial medial meniscectomy (APMM) to differentiate, prior to surgery, between patients likely to benefit from APMM and those at risk for developing medial post-meniscectomy syndrome (MPMS). By integrating knee bone morphology, clinical, and demographic data into an end-to-end machine learning pipeline, we aim to optimize patient selection for APMM. This data-driven approach has the potential to improve surgical success rates and reduce the occurrence of MPMS by identifying patients who may benefit from alternative treatment options.

Samenvatting

Meerdere risicofactoren die bijdragen aan het ontstaan en de progressie van artrose aan de knie zijn al beschreven in de huidige literatuur, waaronder het verlies van een functionele meniscus. Echter, niet alle gevallen van vroegtijdige knie artrose kunnen worden verklaard door de momenteel bekende risicofactoren. Op basis van empirische observaties in de klinische praktijk groeide de hypothese dat de botmorfologie van de knie een rol kan spelen in deze multifactoriële degeneratieve aandoening. Innovaties in biomedische beeldvorming en aritificiële intelligentie ondersteunen de automatisering en schaalbaarheid van 3D bot morfologie analyse. Bovendien maken geautomatiseerde methoden voor het aanduiden van anatomische oriëntatiepunten een nauwkeurige en transparante morfometrische analyse van de knie mogelijk, zelfs in grote datasets. Unieke morfologische kenmerken, zoals een smalle mediale femurcondyle, blijken verband te houden met een degeneratief mediaal knie compartiment. Daarnaast werden in de context van meniscus transplantaties verschillen in grootte opgemerkt tussen de donors (grotere mediale meniscus) en de acceptors (kleinere mediale meniscus). Statistische vormanalyse maakt het mogelijk om de voornaamste variaties in knie vormen te vergelijken tussen patiënten die goed reageren op arthroscopische partiële mediale meniscectomie (APMM) en patiënten met het mediale post-meniscectomiesyndroom (MPMS; klinische diagnose, gekarakteriseerd door terugkerende pijn gelinkt aan het gebrek aan functionele meniscus).

In dit doctoraatsonderzoek werd een prognostisch model ontwikkeld voor APMM om voorafgaand aan de operatie een onderscheid te maken tussen patiënten die waarschijnlijk baat hebben bij APMM en patiënten die risico lopen op het ontwikkelen van het MPMS. De bot morfologie van de knie, klinische en demografische gegevens werden geïntegreerd in een machine learning algoritme, met als doel om de patiëntenselectie voor APMM te optimaliseren. Deze data-gedreven benadering heeft het potentieel om de slaagkans van de operatie te verbeteren en de incidentie van MPMS te verminderen door patiënten te identificeren die waarschijnlijk baat hebben bij alternatieve behandelopties.

Part I

Clinical background and research objectives

Chapter 1

Anatomy of the knee

Contrary to what one might think, the knee joint (and not the hip joint) is the largest joint in the human body. It bears most of our weight and enables essential movement patterns of the lower limb like walking, running and jumping. In this chapter, the necessary anatomical background is provided to better understand the process of knee degeneration described in the next chapter.

1.1 Anatomical reference frame

Firstly, for an accurate description of the human knee anatomy, a set of unambiguous positional terms are needed to describe relative positions of anatomical structures. These always refer to the standard anatomical position: the human body is standing upright and at rest. The three main anatomical planes (Figure 1.1) are the coronal plane (or frontal plane), sagittal plane and axial plane (or transverse plane). A coronal plane divides the body in an anterior (front) and posterior (back) part. Sagittal planes divide a left from a right part and are perpendicular to the mediolateral axis. Medial refers to a position closer to the midline, while lateral means farther away from the midline. Finally, axial planes cut the human body in a superior (upper) and an inferior (lower) part. All three anatomical planes are perpendicular to each other.

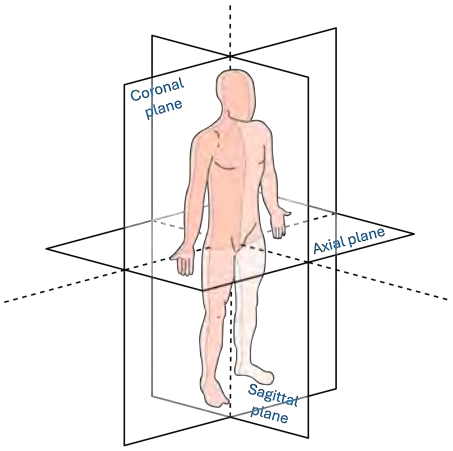


Figure 1.1: The three types of anatomical planes: coronal (or frontal), sagittal and axial (or transverse) planes [1].

1.2 Bony anatomy of the knee

The knee consists of two joints: the tibiofemoral joint is mostly a hinge joint, connecting the femur (thigh bone) and tibia (shin bone) while the patellofemoral joint is a saddle joint, allowing the patella (kneecap) to slide in the trochlear groove of the femur. Acting as a lever arm for the M. Quadriceps and patellar tendon, the patella is the largest sesamoid bone (= bone embedded in tendon or muscle, to increase leverage and reduce friction) in the human body [2] (Figure 1.2). This result in a larger lever arm for the Musculus (M.) Quadriceps and patellar tendon, helping to achieve the same rotational momentum around the flexion-extension axis with a smaller muscle contraction. In this PhD dissertation, we will focus mainly on the tibiofemoral joint, as this is the main load-bearing part of the knee.

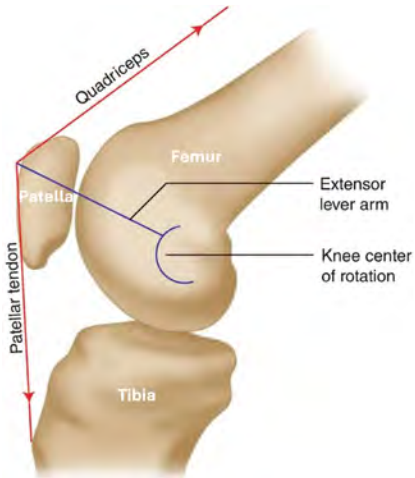


Figure 1.2: Medial view of a right knee: the patella acts as a fulcrum to increase the lever arm of the M. Quadriceps and patellar tendon. Figure adapted from [2].

From a distal view, the distal femur consists of a medial and lateral condyle, and the intercondylar notch in between (Figure 1.3). In a sagittal cross-section, the condyles have a J-curve shape (Figure 1.4), and they articulate with the tibial plateau, allowing for flexion and extension of the knee.

The proximal tibia consists of a medial and lateral tibial plateau, with the intercondylar eminence in between (Figure 1.5).The condylus medialis or medial tibial plateau is part of the medial compartment of the tibiofemoral joint. Similarly, the condylus lateralis or lateral tibial plateau is part of the lateral compartment of the tibiofemoral joint. The eminentia intercondylaris or intercondylar eminence is located between the medial and lateral tibial plateau and is characterized by two bone ridges or prominences mediolaterally: the medial tibial spine at the lateral border of the medial tibial plateau and the lateral tibial spine at the medial border of the lateral tibial plateau.

Figure 1.3: Anatomy of the distal femur. The facies patellaris refers to the trochlear groove and is part of the patellofemoral joint. The medial femoral condyle (or condylus medialis) is part of the medial compartment of the tibiofemoral joint. Similarly, the lateral femoral condyle (or condylus lateralis) is part of the lateral compartment of the tibiofemoral joint. The fossa intercondylaris is also referred to as the femoral notch, the space between the medial and lateral femoral condyle. Figure adapted from [3].

Figure 1.4: Lateral X-ray of a knee. The medial and lateral condyle are in superposition as they are behind one another from a lateral viewpoint. The red J-curve is the contour of the lateral femoral condyle projection.

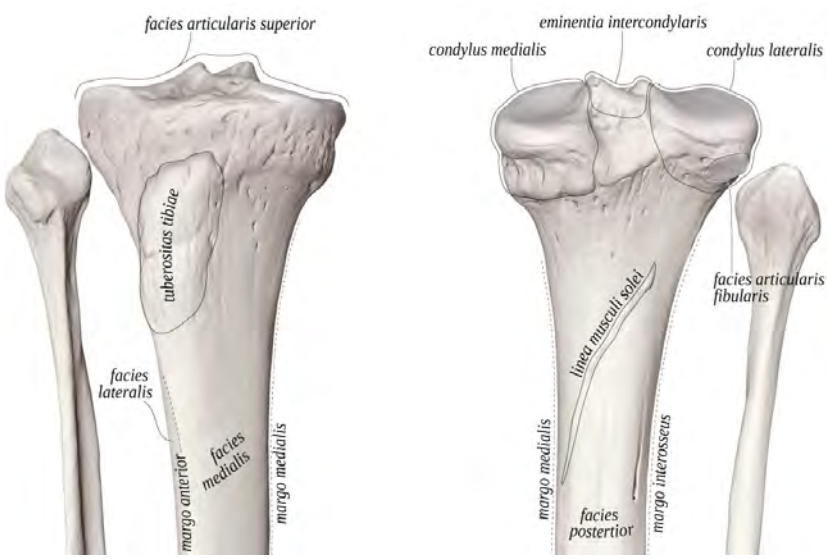


Figure 1.5: Anatomy of the proximal tibia (shin bone) and fibula (calf bone). Figure adapted from [4].

1.2.1 Articular cartilage

The articular contact surfaces of the tibiofemoral joint (medial condyle and tibial plateau, lateral condyle and tibial plateau) are covered with a layer of hyaline cartilage, as well as the articular contact surfaces of the patellofemoral joint (femoral trochlea and patellar facets). The cartilage of the knee is mainly composed of an extracellular matrix rich in collagen fibers and proteoglycans [5]. This specific tissue composition is associated to its functions The collagen fibers ensure tensile strength, enabling it to withstand the imposed stress by weight-bearing activities. Meanwhile, the proteoglycans, with their ability to attract and retain water, create a cushioning effect, allowing for shock-absorption and resistance against compression. In collaboration with the intra-articular synovial fluid, it helps to minimize friction between the two bones during movement. This combination of properties makes knee cartilage an essential component for smooth, pain-free articulation. Articular chondrocytes are the cells that are responsible for this extracellular matrix composition, production and maintenance of the cartilage [6] trough different pathways, mediated by enzymes, growth factors and inflammatory cytokines.

1.3 Soft-tissue enveloppe

1.3.1 Knee capsule

The knee is a large synovial joint, characterized by a layer of cartilage at the femoral and the tibial epiphysis. The articular joint space (between both bone ends) is encapsulated by a fibrous capsule, lined with a layer of synovium at the inside [7]. The synovial cells in the synovium layer produce the synovial fluid, containing collagen, hyaluronan and inflammatory mediators. This viscous synovial fluid acts as a lubricant and nutrient for the adjacent cartilage layer: a very thin layer of synovial fluid between the femoral and tibial cartilage minimizes the friction during flexion and extension of the knee.

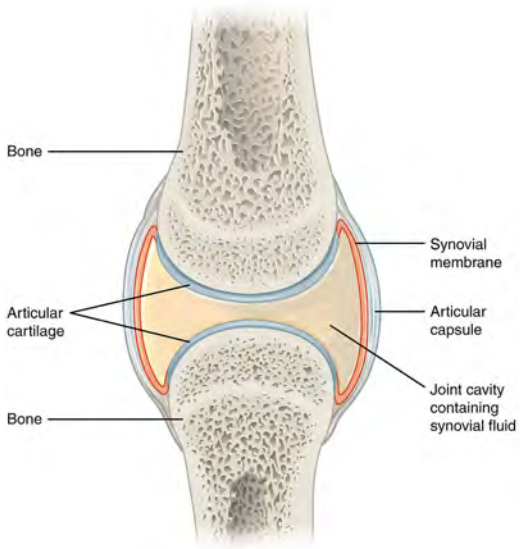


Figure 1.6: Generalization of a human synovial joint. The joint capsule encapsulates both bone ends and is lined with a layer of synovial membrane at the inside [8].

1.3.2 Muscles

M. Quadriceps The M. Quadriceps consists of four separate parts (called 'heads'): three superficial (vastus medialis, laterals and rectus femoris) and one deep (vastus intermedius) head. It is a biarticular muscle, spanning both the hip (however only the rectus femoris head) and the knee anteriorly, thereby acting as a hip flexor (only the rectus femoris) and knee extensor. The vastus medialis originates from the medial side of the femur and inserts together with the tendons of the vastus lateralis, intermedius and rectus femoris at the quadriceps tendon on the patella. The vastus lateralis originates from the greater trochanter, intertrochanteric line and linea aspera of the femur. The vastus intermedius originates from the anterolateral side of the upper two-thirds of the femur. Finally, the rectus femoris originates from the anterior inferior iliac spine and a groove above the rim of the acetabulum. All four heads insert into the quadriceps tendon on the patella, which in turn is linked to the tibia trough patellar tendon at the tibial tuberositas.

Hamstring muscles The hamstring muscles are a muscle group located at the posterior aspect of the thigh. They consist of the M. semitendinosus, the M. semimembranosus and M. biceps femoris. Following their origin at the posterior aspect of the femur and the ischium (hip bone), and their insertion at the posterior part of the proximal tibia, they are the main flexors of the knee.

M. Gastrocnemius The M. Gastrocnemius is also a biarticular muscle, spanning both the knee and the ankle joint, thereby fulfilling a dual action: knee flexion and ankle

plantarflexion. It originates superior to the femur condyles and merges together with the M. Soleus insertion in the Achilles tendon at the calcaneus (heel bone) in the foot.

1.3.3 Ligaments

Cruciate ligaments There are two cruciate ligaments in the knee: the anterior cruciate ligament and the posterior cruciate ligament. The anterior cruciate ligament restricts anterior movement of the tibia to the femur, while the posterior cruciate ligament limits posterior motion of the tibia to the femur [9]. Together, the cruciate ligaments are the primary anteroposterior stabilizers of the knee. Moreover, as secondary stabilizers, they limit the endoration of the tibia w.r.t. the femur.

Collateral ligaments There are two collateral ligaments in the knee: a medial and a lateral collateral ligament (Figure: 1.7). The medial collateral ligament (MCL) is composed of two parts: a superficial MCL (sMCL) and a deep MCL (dMCL) [9]. The sMCL is a broad, flat band that originates from the adductor tubercle on the medial femoral condyle and inserts at the medial tibial metaphysis 4-5 cm below the joint line. The dMCL is attaches the medial meniscus to the distal femur (meniscofemoral ligament) and the proximal tibia (meniscotibial ligament) along the fibers of the joint capsule. Because of its origin and insertion, the complete MCL acts as a stabilizer against valgus forces, which push the knee inward. The LCL originates from the lateral epicondyle and inserts onto the head of the fibula, leading to a resistance against varus forces, which push the knee outward.

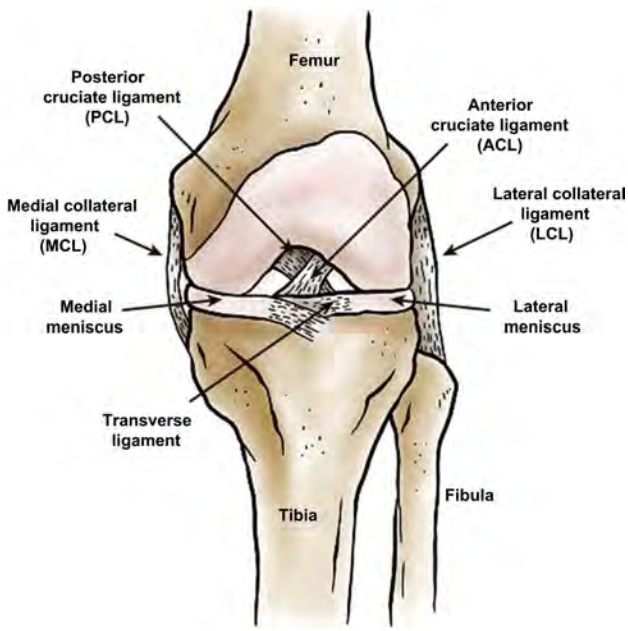


Figure 1.7: Schematic overview of the knee ligaments and menisci in a frontal view. [10]

1.3.4 Menisci

Medial meniscus The medial meniscus is a C-shaped, fibrocartilaginous structure, between the medial femoral condyle and the medial tibial plateau [11]. Its primary function is load distribution, shock absorption, and stabilization of the knee. Anatomically, the medial meniscus is attached firmly to the tibia via the anterior and posterior horns (Figure: 1.8). The peripheral border is thicker than the inner edge, resulting in a typical triangular-like cross-section as seen on MRI. Because of its shape, it enhances the congruency between the convex medial femoral condyle and the rather flat tibial plateau, serving as a knee stabilizer. The meniscus is partially vascularized, with the outer one-third receiving blood supply, contributing to its limited healing capacity. The inner two-thirds, being avascular, rely on synovial fluid for nutrient diffusion and have none to very limited regenerative capacities. The medial meniscus is also anchored to the joint capsule and the medial collateral ligament (dMCL), restricting its mobility compared to the lateral meniscus, and making it more susceptible to injury.

Lateral meniscus The lateral meniscus is an O-shaped, fibrocartilaginous structure, situated between the lateral femoral condyle and the lateral tibial plateau [11]. Similarly to the medial meniscus, it plays a role in load distribution, shock absorption, and joint stability. Unlike the medial meniscus, the lateral meniscus is more uniformly shaped and covers a larger portion of the tibial articular surface. It is attached to the tibia via the anterior and posterior horns (Figure: 1.8) and is also connected to the femur through the meniscofemoral ligaments (Humphrey and Wrisberg ligaments). The lateral meniscus is more mobile due to its looser attachment to the joint capsule and the absence of a firm connection to the lateral collateral ligament. This increased mobility decreases its susceptibility to injury compared to the medial meniscus. Comparable to the medial meniscus, the peripheral one-third of the lateral meniscus is vascularized, aiding in potential healing, while the inner two-thirds are avascular, relying on synovial fluid for nutrient supply. These inner two-thirds therefore also have a limited regeneration capacity.

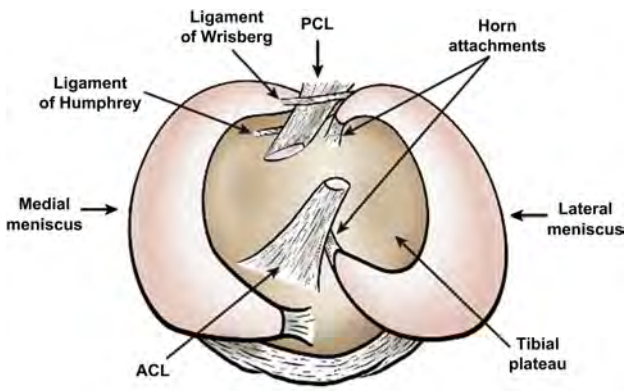


Figure 1.8: Medial and lateral meniscus anatomy in relation to the tibial plateau (top view) and cruciate ligaments. ACL: anterior cruciate ligament, PCL: posterior cruciate ligament. Figure adapted from [10].

1.4 Elementary knee biomechanics

1.4.1 Kinematics of the knee

While the main movement of the knee is flexion and extension, in total the tibia has 6 degrees of freedom to move around the femur: 3 translations and 3 rotations. The final range of motion for each degree of freedom is determined by the complex interplay of osseous geometry, knee ligaments and muscles [12] . During flexion and extension (= the main degree of freedom of the knee), the femoral condyles roll and slide over the tibial plateau. This motion is facilitated by the unique asymmetrical shape of the femoral condyles and the corresponding tibial surfaces, with the medial and lateral menisci acting as congruent, load-distributing intermediaries. During knee flexion, the femur partly translates posterior (Figure 1.9). The anterior and posterior cruciate ligaments (ACL and PCL) play crucial roles in guiding these movements, preventing excessive anterior or posterior tibial translation, respectively. From a certain amount of tension in the ACL (because of the 'backward' rolling of the femur during flexion of the knee), the femur stops rolling posteriorly and continues to rotate further while sliding over the tibial plateau. Therefore, the ACL guides the femoral rollback during the whole range of motion. Note that the lateral femoral condyle typically translates more posteriorly than the medial femoral condyle, leading to some external femoral rotation during flexion of the knee. This can be explained the following mechanisms. Firstly, the lateral meniscus has a greater mobility in comparison to the medial meniscus leading to less motion restriction in the lateral compartment. Additionally, the asymmetrical geometry of distal femur and proximal tibia contributes to a greater posterior rollback in the lateral compartment.

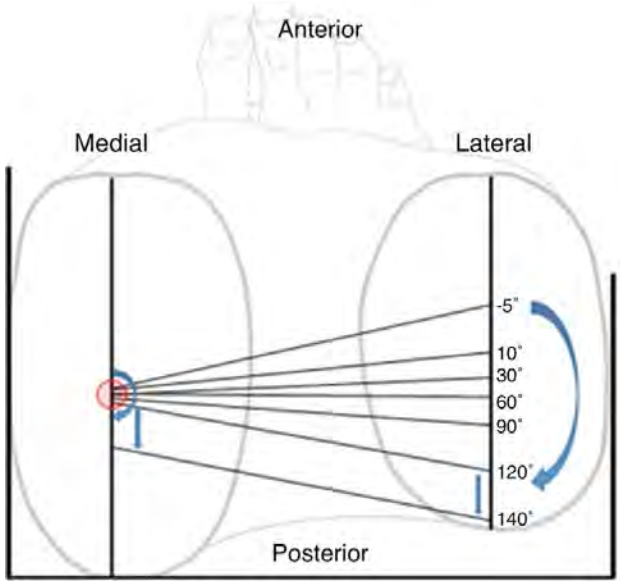


Figure 1.9: Top view of a right tibial plateau with the transepicondylar axis (connecting the medial and lateral epicondyle) projected on it. The arrow indicates the direction of the consequent transepicondylar axis positions on discrete positions trough the flexion range of motion. Figure adapted from [12].

1.4.2 Coronal alignment of the lower limb

Coronal alignment deviations at the level of the knee joint are a known risk factor for unicompartmental knee degeneration, including degenerative meniscal lesions. Three main categories of coronal alignment are considered: varus (bow-legged), neutral and valgus (knock-kneed) alignment (Figure 1.10). In-silico and in-vitro experiments have already established an altered load distribution across the different compartments with varus knees more prone to overload in the medial compartment and valgus knees more prone to overload in the lateral compartment. [13, 14].

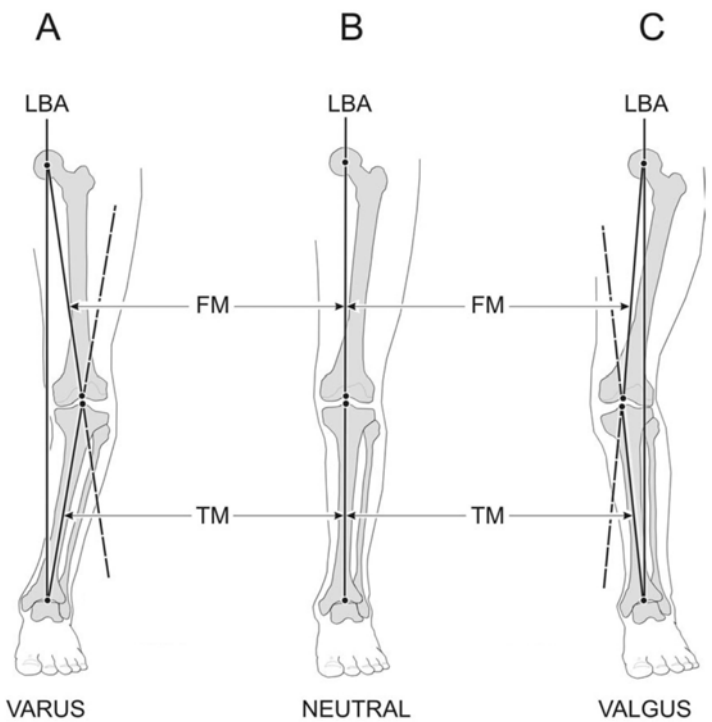


Figure 1.10: The three main categories of coronal knee alignment: (A) varus, (B) neutral and (C) valgus knees. LBA: load-bearing axis (also referenced to as Mikulicz line), FM: femoral mechanical axis, TM: tibial mechanical axis. Figure adapted from [15].

As a rule of thumb, while performing corrective osteotomies for excessive varus or valgus alignment, the surgeon often aims for the line between the center of the hip and the center of the ankle (load-bearing axis (LBA), also referenced to as Mikulicz line) to cross the tibial plateau 4 +/- 2 mm medial to the center of the knee [16]. Additional measurements are also commonly used to assess potential alignment abberrations in the coronal plane (see Figure 1.11), such as the hip-knee-ankle (HKA) angle, femoral mechanical angle (FMA) and tibial mechanical angle (TMA). These measurements enable a more refined classification of alignment abnormalities [17, 18] and serve as indispensable tools for a more patient-specific surgical planning.

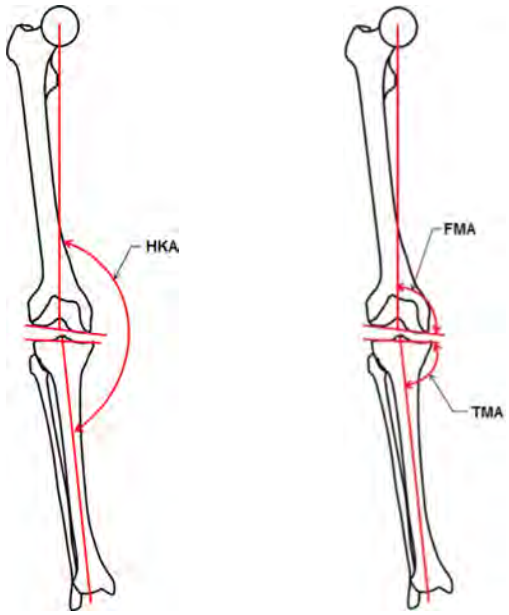


Figure 1.11: (Left) Measurement of the hip-knee-ankle angle (HKA): the angle between the mechanical axes of the femur and tibia in the coronal plane. **(Right)** Measurements of the femoral mechanical angle (FMA) and tibial mechanical angle (TMA). The FMA is the angle as measured from the medial side between the femoral mechanical axis and a tangent to the distal femoral condyles in the coronal plane. The TMA is then the angle, again measured from the medial side, between the tibial mechanical axis and a tangent to the tibial plateau in the coronal plane [19, 20].

Bibliography

[1] CFCF. File:Anatomical Planes.svg — Wikimedia Commons, the free media repository; 2023. [Online; accessed 28-July-2024]. Available from: `https://commons.wikimedia.org/w/index.php?title=File:Anatomical_Planes.svg&oldid=791936127`.

[2] Reynolds RJ, Michelet A, Müller JH, Saffarini M. In: Kinematics of the Native Knee. Springer International Publishing; 2022. p. 19-42.

[3] Šavlovskis J, Raits K. Femur (Thigh bone); 2020. [Online; accessed 28-July-2024]. Available from: `https://www.anatomystandard.com/ossa-et-juncturae/extremitas-inferior/femur.html`.

[4] Šavlovskis J, Raits K. Tibia (Shin bone); 2020. [Online; accessed 28-July-2024]. Available from: `https://www.anatomystandard.com/ossa-et-juncturae/extremitas-inferior/tibia.html`.

[5] Fox AJS, Bedi A, Rodeo SA. The Basic Science of Articular Cartilage: Structure, Composition, and Function. Sports Health: A Multidisciplinary Approach. 2009 11;1:461-8.

[6] Carballo CB, Nakagawa Y, Sekiya I, Rodeo SA. Basic Science of Articular Cartilage. Clinics in Sports Medicine. 2017;36(3):413-25. Articular Cartilage.

[7] Flandry F, Hommel G. Normal Anatomy and Biomechanics of the Knee. Sports Medicine and Arthroscopy Review. 2011 6;19:82-92.

[8] College O. File:907 Synovial Joints.jpg — Wikimedia Commons, the free media repository; 2023. [Online; accessed 28-July-2024]. Available from: https://commons.wikimedia.org/w/index.php?title=File:907_Synovial_Joints.jpg&oldid=771472153.

[9] Hassebrock JD, Gulbrandsen MT, Asprey WL, Makovicka JL, Chhabra A. Knee Ligament Anatomy and Biomechanics. Sports Medicine and Arthroscopy Review. 2020 9;28:80-6.

[10] Makris EA, Hadidi P, Athanasiou KA. The knee meniscus: structure-function, pathophysiology, current repair techniques, and prospects for regeneration. Biomaterials. 2011 10;32:7411-31.

[11] Fox AJS, Wanivenhaus F, Burge AJ, Warren RF, Rodeo SA. The human meniscus: A review of anatomy, function, injury, and advances in treatment. Clinical Anatomy. 2015 3;28:269-87.

[12] Pinskerova V, Vavrik P. In: Charles PAR, Vendittoli, editors. Knee Anatomy and Biomechanics and its Relevance to Knee Replacement. Springer International Publishing; 2020. p. 159-68.

[13] Willinger L, Lang JJ, Berthold D, Muench LN, Achtnich A, Forkel P, et al. Varus alignment aggravates tibiofemoral contact pressure rise after sequential medial meniscus resection. Knee Surgery, Sports Traumatology, Arthroscopy. 2020 4;28:1055-63.

[14] Van Oevelen A, den Borre IV, Duquesne K, Pizurica A, Victor J, Nauwelaers N, et al. Wear patterns in knee OA correlate with native limb geometry. Frontiers in bioengineering and biotechnology. 2022;10:1042441.

[15] Cooke TDV, Sled EA, Scudamore RA. Frontal plane knee alignment: a call for standardized measurement. The Journal of rheumatology. 2007 9;34:1796-801.

[16] Luís NM, Varatojo R. Radiological assessment of lower limb alignment. EFORT Open Reviews. 2021 6;6:487-94.

[17] Hirschmann MT, Moser LB, Amsler F, Behrend H, Leclerq V, Hess S. Functional knee phenotypes: a novel classification for phenotyping the coronal lower limb alignment based on the native alignment in young non-osteoarthritic patients. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 5;27:1394-402.

[18] MacDessi SJ, Griffiths-Jones W, Harris IA, Bellemans J, Chen DB. Coronal Plane Alignment of the Knee (CPAK) classification. The Bone & Joint Journal. 2021 2;103-B:329-37.

[19] Hirschmann MT, Hess S, Behrend H, Amsler F, Leclercq V, Moser LB. Phenotyping of hip–knee–ankle angle in young non-osteoarthritic knees provides better understanding of native alignment variability. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 5;27:1378-84.

[20] Hirschmann MT, Moser LB, Amsler F, Behrend H, Leclercq V, Hess S. Phenotyping the knee in young non-osteoarthritic knees shows a wide distribution of femoral and tibial coronal alignment. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 5;27:1385-93.

Chapter 2

Degenerative meniscus and knee osteoarthritis

Building further on the essentials of knee anatomy as provided in the previous chapter, in this chapter the clinical challenge is described and situated within the degenerative knee spectrum. The focus lies on the medial post-menisectomy syndrome, that eventually leads to accelerated knee osteoarthritis. Given its large incidence, and for matters of completeness, regular (longer term) post-meniscectomy knee osteoarthritis is also discussed.

2.1 Introduction

Degenerative meniscus tears are strongly associated with knee osteoarthritis; a chronic condition characterized by the progressive degradation of articular cartilage within the knee joint. This deterioration manifests clinically as inflammation, pain, stiffness, and functional impairment. The precise etiological relationship between degeneration of the meniscus and knee osteoarthritis remains an area of active investigation. Two primary hypotheses have been proposed [1, 2]:

1. The absence of a structurally intact meniscus may predispose the knee to osteoarthritic changes due to altered load bearing and distribution.
2. Both cartilage and menisci are subject to the same inflammatory environment (promoting extracellular matrix destruction) and pathological load distributions (e.g. excessive varus or valgus knees) that are associated with knee osteoarthritis.

While the current standard treatment for meniscal lesions aims to preserve as much of the meniscal tissue as possible, a total meniscectomy was one of the most frequently performed knee surgeries in the previous decades. Still today, many of these patients present in clinical practice with the consequences of a meniscus-deficient knee. As already described in 1948 (!) by Fairbank et al. [3], radiographic joint space narrowing (due to cartilage wear, radiographs show bone on bone in already advanced cases) and degenerative changes occur more frequently in meniscus-deficient knees in comparison

to those with an intact meniscus [4, 5]. While arthroscopic techniques and optimized meniscal tissue-sparing approaches have been refined over time, irreparable meniscal injuries continue to necessitate meniscectomy, leading to a functional meniscal deficiency.

2.2 Degenerative meniscus lesions

2.2.1 Epidemiology and pathophysiology

Meniscus tears are primarily categorized based on their origin: traumatic or non-traumatic. Although these two pathogenic pathways share some risk factors such as lower limb coronal malalignment (excessive varus or valgus) and morbid obesity, the mechanism of action in their pathogenesis is almost completely different. Therefore, they should be considered as two distinct pathologies when considering treatment options.

In contrast with traumatic lesions (as a result of knee trauma, predominantly in younger active individuals, typically radial tears), degenerative meniscus lesions (Figure 2.1) are generally non-traumatic and a part of normal aging. The ESSKA European meniscus consensus group defined meniscus degeneration generally as follows: a slowly developing process typically involving a horizontal cleavage in the meniscus of a middle-aged or older person [6]. Their prevalence increases with increasing patient age: from 19% in the group of 50-59 years old women to over 50% in the group of men between 70 and 90 years old [7]. Their typology is often horizontal, (oblique) flap, or complex tears (combination of different characteristics), and they are associated with older age and no history of recent knee trauma. They typically occur at the body or posterior horn of the medial meniscus. Figure 2.2 gives an visual overview of some of the most frequent meniscal tear patterns.

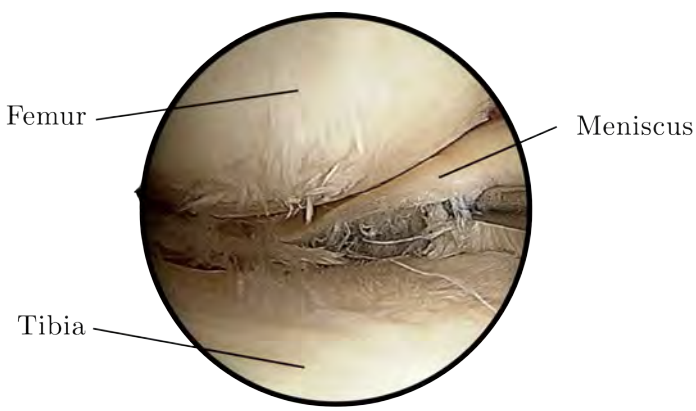


Figure 2.1: Arthroscopy image of a degenerative medial meniscus (posterior horn). Note the ragged, torn edges of the meniscus. Image adapted from [8].

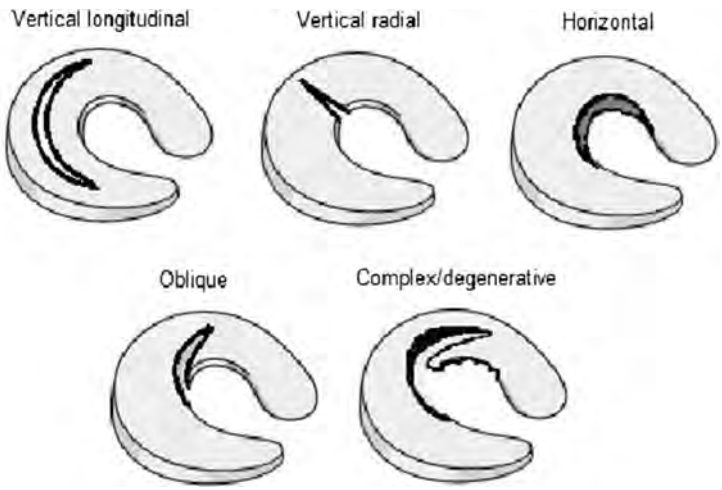


Figure 2.2: Simplified visual representation for some of the most frequent meniscus tear patterns. [9].

2.2.2 Treatment options for degenerative meniscus lesions

Arthroscopic partial meniscectomy has been the standard treatment of (degenerative) meniscal tears for many years now. To date, it is still one of the most frequently performed knee surgeries worldwide [10, 11, 12, 13, 14]. It involves removing the damaged portions of the meniscus during arthroscopy, a minimally invasive technique with tiny incisions and a camera to visualize the intra-articular joint space. This procedure aims to alleviate pain and enhance joint function by eliminating unstable meniscal fragments that cause the mechanical symptoms. However, as it became clear that arthroscopic partial meniscectomy is associated with a significant greater risk to develop knee osteoarthritis [15, 16], its indications were scrutinized and there has been a paradigm shift towards saving the meniscal tissue as much as possible [17].

Current guidelines recommend a conservative, non-surgical approach as the initial management strategy for symptomatic degenerative tears. Incidental findings without associated symptoms do not require intervention. Conservative treatment includes pharmacological treatment (analgesics and non-steroidal anti-inflammatory drugs) to alleviate pain and inflammation, as well as physiotherapy to strengthen the periarticular knee musculature and enhance knee stability. According to the ESSKA guidelines for surgical treatment of degenerative meniscal tears [6], a time interval of at least 3 months after the initial symptom onset should be considered before the decision of surgical treatment is made. Surgery might be indicated earlier in case of considerable mechanical symptoms (e.g. restricted range of motion, 'locking' of the knee,...). Degenerative meniscal lesions are typically not amendable to meniscus repair (sutures). The nature of these lesions, often characterized by complex patterns, and its location typically within the avascular portion of the meniscus, severely limit the tissue's inherent healing capacity. Moreover, the degenerative process often compromises the structural integrity of the meniscal tissue, making it suboptimal for suture fixation. These factors collectively contribute to an elevated risk of repair failure. Therefore,

because meniscal repair (with sutures) is typically unfeasible in these cases, arthroscopic partial meniscectomy remains a valid treatment option with the goal of immediate pain relief while preserving as much meniscal tissue as possible.

2.3 Post-meniscectomy syndrome

2.3.1 Epidemiology and pathophysiology

In addition to the majority of patients going towards end stage knee OA in the course of several years (typically more than two years), a significant proportion of the meniscectomy patients (between 6 and 25% [18]) will develop the post-meniscectomy (pain) syndrome (Figure 2.3). This condition adds to the burden of the normally expected knee osteoarthritis, which affects a majority of patients over time (rather long-term consequence). It is a clinical complication after (partial) meniscectomy, characterized by a toothache-like dull and nagging pain in combination with transient joint effusions, after a short pain-free interval (typically within two years post-meniscectomy), with or without additional cartilage damage [19]. The removal of the meniscus, a critical load-bearing and shock-absorbing structure, alters the biomechanical environment of the knee joint, resulting in increased mechanical stress on the articular cartilage and subchondral bone. These altered biomechanics, combined with the inflammatory environment associated with the meniscal injury and surgical manipulation, is thought to accelerate the degenerative process [20], finally culminating in pain, further meniscal tissue degradation and accelerated progression of knee osteoarthritis [21, 22]. Given the prevalence of the (partial) meniscectomy procedure worldwide, even in the face of meniscus-saving guidelines, this patient subgroup cannot be overlooked. For those with unrelieved pain after conservative treatment and who are ineligible for meniscus repair, the partial meniscectomy remains a valuable treatment option.

(a)

(b)

Figure 2.3: Potential consequences of a functional meniscus deficiency. **(a)** Average situation post-meniscectomy: the knee evolves gradually to end-stage knee OA in an unknown time (but generally speaken more than 5 years). **(b)** The post-meniscectomy syndrome is associated with accelerated knee degeneration, leading to early end-stage knee OA.

2.3.2 Treatment options for post-meniscectomy syndrome

Conservative treatment is the first step in the treatment algorithm of the post-meniscectomy syndrome [19]. The meniscectomy resulted in a significant change of load distribution in the tibiofemoral joint. Therefore, conservative treatment includes the unloading of the affected knee (compartment) as much as possible by a reduction in activity or sports. Furthermore, the use of crutches or unloader braces, insoles and weight loss are additional tools to achieve this goal [23]. Personalized physiotherapy is another conservative treatment option to improve quadriceps strength and core control, potentially leading to a better control of the symptoms [24]. Currently, there is no consensus to include intra-articular injection therapy, either with corticosteroids [19], hyaluronic acid [25], platelet-rich plasma (PRP) [26] or mesenchymal stem cells [27].

When non-surgical treatment fails, surgical management of mechanical risk factors is the next step to restore the knee 'homeostasis' [28]. Following factors should be assessed and corrected if necessary: lower limb alignment, knee stability and meniscal deficiency. Lower limb malalignment can be corrected by performing osteotomy surgery, either at the level of the distal femur (less frequent) and/or proximal tibia, depending on the specific origin of the malalignment. In addition, knee instability originating from ligament lesions or insufficiency should be managed by ligament

reconstruction, with a special attention for ACL insufficiency [29]. After the diagnosis of a symptomatic meniscal deficiency (root tear or removal of significant amount of meniscal tissue) in a well-aligned and stable knee, meniscal substitution therapy can be considered; including meniscus allograft transplantation (MAT), a meniscal scaffold or even an artificial meniscus implant.

Meniscus scaffold

Meniscal scaffolds are three-dimensional biocompatible structures, capable of supporting meniscus-like fibrocartilaginous tissue regeneration. They are well-suited to treat partial or segmental meniscus defects [28]. The ideal patient for this treatment option still has an intact meniscal rim, including intact anterior and posterior meniscal horns. This ensures the scaffold can be fixated properly. A BMI below 35 kg/m^2 and cartilage damage below ICRS grade 3a are then the final additional patient selection criteria [30]. The aim of a meniscal scaffold is to allow meniscus-like tissue to grow into the scaffold and consequently restore and maintain the original meniscal function [31].

Meniscus allograft transplantation

When meniscal scaffolds are no option (peripheral meniscal rim transsected, meniscal root tears or a large amount of meniscal tissue got lost) a meniscus allograft transplantation (MAT) is an option to substitute the native meniscus. It has been proven to be efficacious both from a biomechanical (reduction of contact stress [32]) and clinical success rate [33] point of view. Given the scarcity in donor menisci, this treatment option is still limited to younger patients (<50-55 years) [34, 18] with a higher rehabilitation potential.

Artificial meniscus implant

Finally, as a last and still more experimental treatment option, an artificial meniscus implant can be used to mimic the function of the natural meniscus [18]. Artificial meniscus implants are synthetic, permanent devices designed to mimic the mechanical function of a natural meniscus and aim to provide long-term relief from the symptoms associated with meniscal loss. At the moment of writing, no such implant is commercially available anymore [35, 20], though several are under development.

Prosthesis surgery

In cases where contraindications exist for the previously discussed meniscal substitution therapies, a (unicompartmental) knee prosthesis is indicated when conservative treatment (including pain management) have failed, surgical management of mechanical risk factors has also failed or the patient has already end-stage knee osteoarthritis (bone on bone, severe osteophytes,...).

2.3.3 Prevention of the post-meniscectomy syndrome

It is clear the first step in preventing the post-meniscectomy syndrome is to save the meniscus to the fullest extent possible. Additionally, analogous steps as in the actual treatment of the post-meniscectomy syndrome might also have a preventive effect, such as unloading the affected knee compartment, as well as ensuring a proper lower limb alignment and aiming for a stable knee.

2.4 Knee osteoarthritis in the meniscus-deficient knee

2.4.1 Epidemiology and pathophysiology

Knee osteoarthritis (OA) in general is a highly prevalent degenerative joint disease marked by the progressive deterioration of articular cartilage, subchondral bone sclerosis, and the formation of osteophytes [36]. It manifests as pain, stiffness, and decreased function in the affected knee, severely impacting the quality of life for those afflicted . The pathophysiology of knee OA is multifaceted, involving both mechanical and biological factors. Abnormal joint loading, genetic predisposition, and inflammatory mediators all play a role in its development and progression.

While the previously described post-meniscectomy syndrome (see 2.3.1: Epidemiology and pathophysiology) is a short-term consequence in the meniscus deficient knee, more than 50% of the meniscectomy patients ultimately develop progressive symptomatic knee osteoarthritis (OA) in the longer term [5, 37] (Figure 2.3). Therefore, knee OA can be seen as a more chronic, slower consequence of the same traumatic or degenerative event (loss of meniscal function [1]) that causes the more acute post-meniscectomy syndrome.

2.4.2 Delay of post-meniscectomy knee osteoarthritis

As the degenerative process of knee osteoarthritis is also (to a certain extent) part of normal aging, the delay and slowing down of its onset and progression are a more feasible aim then full prevention [36]. Obviously, meniscal substitution strategies as delay strategy for knee osteoarthritis (OA) only make sense when there is no knee OA yet and the cartilage damage is still limited. Unfortunately, at the time of writing, meniscal substitution strategies cannot yet be implemented into routine clinical care for every patient. The supply of matching healthy menisci from biobanks is scarce (see also chapter 3: Meniscus Size Differs Between Patient and Donor Populations for Meniscus Allograft Transplantation), artificial implants currently exhibit a relatively high failure rate [38] and long-term clinical follow-up data remains limited [39].

Therefore, anticipating approaches for post-meniscectomy knee OA remain restricted to the assessment of known mechanical risk factors such as coronal malalignment of the lower limb, knee instability or excessive knee loading (either by activity level, high-impact sports or overweight), analoguous to the workup in the treatment of the post-meniscectomy syndrome [28]. Indeed, the treatment of the post-meniscectomy syndrome also aims to slow down the progression of knee OA . This can be done

through a combination of conservative (lifestyle changes, pain management, unloader brace, physiotherapy,...) and surgical interventions (osteotomy, ligament reconstruction, focal cartilage defect repair). In the last few decades, there has been a paradigm shift towards joint-sparing surgery whenever possible [40]. Osteotomies, which involve cutting and realigning bones to redistribute mechanical load and relieve pressure on the affected area, have gained popularity. This surgical intervention can delay the need for joint replacement and preserve the patient's native joint for a longer period. Other joint-sparing procedures, such as cartilage repair techniques for small focal lesions including microfracture, autologous chondrocyte implantation (ACI), and osteochondral autograft transplantation (OAT), aim to restore damaged cartilage and improve joint function, albeit without achieving complete regeneration.

This paradigm shift towards joint preservation is also reflected in the development of biological therapies [41]. Treatments such as platelet-rich plasma (PRP) and mesenchymal stem cell (MSC) injections are being investigated for their potential to promote tissue healing and modulate inflammation [42, 43]. Although these therapies are still under study and not universally accepted as standard care, they represent a promising direction in the effort to preserve joint health and function. Biologicals (-mabs), targeting specific inflammatory pathways, are also under research for their potential to modify disease progression [43].

2.4.3 Treatment options for post-meniscectomy knee osteoarthritis

Despite years of research by rheumatologist and the pharma-industry, no effective disease-modifying osteoarthritis drugs are commercialized on the market yet [41]. Additionally, no orthopedic procedures allow for the regeneration of the degenerated cartilage [36]. The management of knee osteoarthritis thus focuses on alleviating symptoms, improving joint function, and slowing down disease progression (see 2.4.2: Delay of post-meniscectomy knee osteoarthritis).

When all previously described joint-sparing management strategies fail, a unicompartmental or total knee arthroplasty (TKA), commonly known as joint replacement surgery, is performed. This procedure involves the surgical removal of the damaged joint surfaces and their replacement with metal and plastic prosthetic components designed to mimic the movement of a healthy knee. The primary goals of TKA are to relieve pain, restore joint function, and improve the patient's overall quality of life. Post-surgical outcomes for TKA are generally positive, with a satisfaction rate of about 80-90% [44, 45], indicating that the majority of patients experience substantial improvements in their symptoms and functional abilities. Factors contributing to patient satisfaction include pain relief, improved mobility, and the ability to return to daily activities.

Overall, the emphasis on joint-sparing interventions underscores the importance of early diagnosis. By prioritizing joint preservation, healthcare providers aim to improve long-term outcomes for patients with knee osteoarthritis, enhancing their quality of life and delaying the need for more invasive procedures such as total knee arthroplasty. Effective management of knee OA requires a personalized approach, taking into account the patient's symptoms, disease severity, and overall health status. This individualized strategy ensures that patients receive the most appropriate and effective

treatments, thereby optimizing outcomes and enhancing quality of life.

2.5 Consolidating mechanical risk factors in knee OA

Articular cartilage is a specialized connective tissue composed mainly of water, collagen (predominantly type II), and proteoglycans, which forms the extracellular matrix (ECM) (cfr. section 1.2.1: Articular cartilage). The ECM provides structural support and resilience, allowing the cartilage to withstand compressive forces. The maintenance of this matrix is controlled by chondrocytes, the only cells residing within cartilage (only 2% of the total tissue volume), which are responsible for balancing the synthesis on one hand and the degradation on the other hand of ECM components [46].

Chondrocytes maintain a balance between synthesis and degradation of ECM by responding to mechanical stimuli through a process called mechanotransduction [47]. This involves converting physical forces into biochemical signals. Key to this process are mechanosensors such as integrins and calcium channels [48, 49], which regulate the cell's response to different levels of strain. Under normal physiological conditions, mechanical stimulation encourages chondrocytes to produce ECM enzymes, facilitating cartilage repair and health. However, as mentioned earlier, under excessive strain, these cells shift towards a catabolic state, releasing other enzymes like matrix metalloproteinases (MMPs) that break down cartilage, contributing to degeneration [50]. While the chondrocytes self are at microscale, mechanical stimuli caused by loading the articular cartilage at macroscale are transduced by the surrounding ECM to the level of the chondrocytes [51].

Thus, the local mechanical stimulus at the level of the chondrocyte is linked to the cartilage tissue stress (and strain: the resulting deformation caused by this amount of stress) at that specific spatial position. Stress is a physical quantity defined as the amount of force exerted per unit area, as experienced by a material. In maximally simplified static knee model, the stress in the cartilage (without making distinction of different layers with different material properties, nor taking anisotropic material characteristics into account) can then be calculated as follows:

$$\sigma = \frac{F}{A} \tag{2.1}$$

where F equals the applied compressive force or load on the cartilage and A equals the contact surface area.

All of the previously described mechanical risk factors for knee OA either contribute to the force or the contact area (Figure 2.4). A key contributor to the force exerted on the knee joint is the patient's body mass, rather than BMI. As the body is subject to gravitational forces ($F_g = m * g$, with F_g the gravitational force, m the patient's body mass and g the gravitational constant) during the stance phase, this force is transmitted to the knee joint. Next, the coronal alignment of the knee (see 1.4.2: Coronal alignment of the lower limb) defines how that weight or load is distributed over the knee compartments and hence contributes also to the force as locally experienced by the cartilage. Finally, also sports and activity level have a multiplicative effect to the load applied to the knee joint, with e.g. running causing a load of up to 4-8 times the body weight on the knee [52].

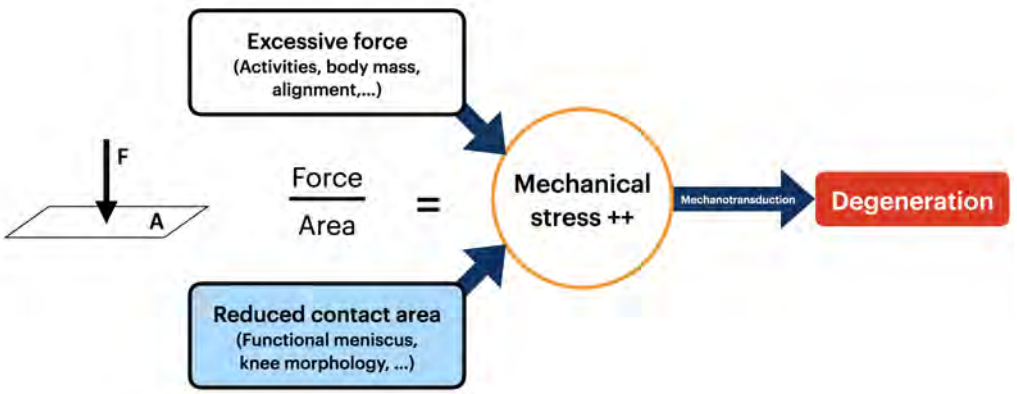


Figure 2.4: Schematic overview of the mechanical risk factors hypothesized to drive knee degeneration. Both the applied load or force on the knee joint, as well as the contact surface (capacity to bear that load) are included.

In previous research, the emphasis has predominantly been on the load exerted on the knee joint, with relatively little attention given to the denominator of the equation: the size of the contact area, which may be significantly affected by inter-individual variability in knee bony morphology, as well as the presence of a functional meniscus.

Although BMI is widely considered a major risk factor for knee osteoarthritis, it is the patient's actual body weight that has the most significant mechanical effect on knee degeneration. The height factor in the BMI formula (i.e., inverse of height squared) has minimal, if any, relevance to knee biomechanics, suggesting that BMI lacks direct relevance in the pathological cascade of knee OA. Despite this, previous studies showing BMI as a significant risk factor remain valid. This apparent paradox can be resolved by interpreting BMI as a surrogate marker, where patient height may indirectly represent the size of the knee joint's contact area. Logically, individuals with smaller statures have smaller knees, and taller individuals have larger knee structures. Therefore also the knee size is hypothesized to be a risk factor in knee OA.

In this work, variations in the knee joint contact area, driven by individual differences in bony morphology, are hypothesized to significantly modulate the impact of body weight on the risk of knee osteoarthritis. Individuals with smaller weight-bearing knee joint contact areas could exhibit a higher susceptibility to knee degeneration due to an increased mechanical stress in the cartilage and menisci, in the absence of other significant risk factors (lower limb malalignment, morbid obesity, meniscal status,...). The following chapter outlines the clinical research objectives, as well as the necessary technological framework needed to address them.

Bibliography

[1] Englund M, Guermazi A, Lohmander SL. The Role of the Meniscus in Knee Osteoarthritis: a Cause or Consequence? Radiologic Clinics of North America. 2009

7;47:703-12.

[2] Edd SN, Giori NJ, Andriacchi TP. The role of inflammation in the initiation of osteoarthritis after meniscal damage. Journal of Biomechanics. 2015 6;48:1420-6.

[3] Fairbank TJ. Knee joint changes after meniscectomy. The Journal of Bone and Joint Surgery British volume. 1948 11;30-B:664-70.

[4] Roos H, Laurén M, Adalberth T, Roos EM, Jonsson K, Lohmander LS. Knee osteoarthritis after meniscectomy: prevalence of radiographic changes after twenty-one years, compared with matched controls. Arthritis and rheumatism. 1998 4;41:687-93.

[5] Roos EM, Ostenberg A, Roos H, Ekdahl C, Lohmander LS. Long-term outcome of meniscectomy: symptoms, function, and performance tests in patients with or without radiographic osteoarthritis compared to matched controls. Osteoarthritis and cartilage. 2001 5;9:316-24.

[6] Beaufils P, Becker R, Kopf S, Englund M, Verdonk R, Ollivier M, et al. Surgical management of degenerative meniscus lesions: the 2016 ESSKA meniscus consensus. Knee Surgery, Sports Traumatology, Arthroscopy. 2017 2;25:335-46.

[7] Englund M, Guermazi A, Gale D, Hunter DJ, Aliabadi P, Clancy M, et al. Incidental Meniscal Findings on Knee MRI in Middle-Aged and Elderly Persons. New England Journal of Medicine. 2008 9;359:1108-15.

[8] Doral MN, Bilge O, Huri G, Turhan E, Verdonk R. Modern treatment of meniscal tears. EFORT Open Reviews. 2018 5;3:260-8.

[9] Karia M, Ghaly Y, Al-Hadithy N, Mordecai S, Gupte C. Current concepts in the techniques, indications and outcomes of meniscal repairs. European Journal of Orthopaedic Surgery & Traumatology. 2019 4;29:509-20.

[10] Abrams GD, Frank RM, Gupta AK, Harris JD, McCormick FM, Cole BJ. Trends in Meniscus Repair and Meniscectomy in the United States, 2005-2011. The American Journal of Sports Medicine. 2013 10;41:2333-9.

[11] Katano H, Koga H, Ozeki N, Otabe K, Mizuno M, Tomita M, et al. Trends in isolated meniscus repair and meniscectomy in Japan, 2011–2016. Journal of Orthopaedic Science. 2018 7;23:676-81.

[12] Thorlund JB, Hare KB, Lohmander LS. Large increase in arthroscopic meniscus surgery in the middle-aged and older population in Denmark from 2000 to 2011. Acta orthopaedica. 2014 6;85:287-92.

[13] Jacquet C, Pujol N, Pauly V, Beaufils P, Ollivier M. Analysis of the trends in arthroscopic meniscectomy and meniscus repair procedures in France from 2005 to 2017. Orthopaedics & Traumatology: Surgery & Research. 2019 6;105:677-82.

[14] Hall MJ, Schwartzman A, Zhang J, Liu X. Ambulatory Surgery Data From Hospitals and Ambulatory Surgery Centers: United States, 2010. National health statistics reports. 2017 2:1-15.

[15] Abram SGF, Judge A, Beard DJ, Carr AJ, Price AJ. Long-term rates of knee arthroplasty in a cohort of 834 393 patients with a history of arthroscopic partial meniscectomy. The bone & joint journal. 2019 9;101-B:1071-80.

[16] Sihvonen R, Paavola M, Malmivaara A, Itälä A, Joukainen A, Kalske J, et al. Arthroscopic partial meniscectomy for a degenerative meniscus tear: A 5 year follow-up of the placebo-surgery controlled FIDELITY (Finnish Degenerative Meniscus Lesion Study) trial. British Journal of Sports Medicine. 2020.

[17] Beaufils P, Pujol N. Management of traumatic meniscal tear and degenerative meniscal lesions. Save the meniscus. Orthopaedics & traumatology, surgery & research : OTSR. 2017 12;103:S237-44.

[18] Zaslav KR, Farr J, Alfred R, Alley RM, Dyle M, Gomoll AH, et al. Treatment of post-meniscectomy knee symptoms with medial meniscus replacement results in greater pain reduction and functional improvement than non-surgical care. Knee Surg Sports Traumatol Arthrosc. 2022 4;30:1325-35.

[19] Drobnič M, Ercin E, Gamelas J, Papacostas ET, Slynarski K, Zdanowicz U, et al. Treatment options for the symptomatic post-meniscectomy knee. Knee Surg Sports Traumatol Arthrosc. 2019 6;27:1817-24.

[20] van Minnen BS, van Tienen TG. The Current State of Meniscus Replacements. Current Reviews in Musculoskeletal Medicine. 2024 5;17:293-302.

[21] Roemer FW, Kwoh CK, Hannon MJ, Hunter DJ, Eckstein F, Grago J, et al. Partial meniscectomy is associated with increased risk of incident radiographic osteoarthritis and worsening cartilage damage in the following year. European Radiology. 2017 1;27:404-13.

[22] Longo UG, Ciuffreda M, Candela V, Rizzello G, D'Andrea V, Mannering N, et al. Knee Osteoarthritis after Arthroscopic Partial Meniscectomy: Prevalence and Progression of Radiographic Changes after 5 to 12 Years Compared with Contralateral Knee. The Journal of Knee Surgery. 2019 5;32:407-13.

[23] Rao AJ, Erickson BJ, Cvetanovich GL, Yanke AB, Bach BR, Cole BJ. The Meniscus-Deficient Knee. Orthopaedic Journal of Sports Medicine. 2015 10;3:232596711561138.

[24] Smith NA, Parsons N, Wright D, Hutchinson C, Metcalfe A, Thompson P, et al. A pilot randomized trial of meniscal allograft transplantation versus personalized physiotherapy for patients with a symptomatic meniscal deficient knee compartment. The Bone & Joint Journal. 2018 1;100-B:56-63.

[25] Filardo G, Di Matteo B, Tentoni F, Cavicchioli A, Di Martino A, Lo Presti M, et al. No effects of early viscosupplementation after arthroscopic partial meniscectomy: a randomized controlled trial. The American Journal of Sports Medicine. 2016;44(12):3119-25.

[26] Zhang JY, Fabricant PD, Ishmael CR, Wang JC, Petrigliano FA, Jones KJ. Utilization of platelet-rich plasma for musculoskeletal injuries: an analysis of current treatment trends in the United States. Orthopaedic journal of sports medicine. 2016;4(12):2325967116676241.

[27] Chew E, Prakash R, Khan W. Mesenchymal stem cells in human meniscal regeneration: a systematic review. Annals of medicine and surgery. 2017;24:3-7.

[28] Bloch B, Getgood A, Parkinson B, Spalding T. In: Concepts in Managing the Patient with Post-meniscectomy Knee Pain. Springer Berlin Heidelberg; 2016. p. 437-46.

[29] Korpershoek JV, de Windt TS, Vonk LA, Krych AJ, Saris DBF. Does Anterior Cruciate Ligament Reconstruction Protect the Meniscus and Its Repair? A Systematic Review. Orthopaedic journal of sports medicine. 2020 7;8:2325967120933895.

[30] Moran CJ, Withers DP, Kurzweil PR, Verdonk PC. Clinical Application of Scaffolds for Partial Meniscus Replacement. Sports Medicine and Arthroscopy Review. 2015 9;23:156-61.

[31] Rodkey WG, DeHaven KE, Montgomery WH, Baker CL, Beck CL, Hormel SE, et al. Comparison of the collagen meniscus implant with partial meniscectomy. A prospective randomized trial. The Journal of bone and joint surgery American volume. 2008 7;90:1413-26.

[32] Alhalki MM, Hull ML, Howell SM. Contact mechanics of the medial tibial plateau after implantation of a medial meniscal allograft. A human cadaveric study. The American journal of sports medicine. 2000;28:370-6.

[33] Romandini I, Grassi A, Lucidi GA, Filardo G, Zaffagnini S. 10-Year Survival and Clinical Improvement of Meniscal Allograft Transplantation in Early to Moderate Knee Osteoarthritis. The American journal of sports medicine. 2024 7;52:1997-2007.

[34] Zaffagnini S, Grassi A, Macchiarola L, Stefanelli F, Coco V, Marcacci M, et al. Meniscal Allograft Transplantation Is an Effective Treatment in Patients Older Than 50 Years but Yields Inferior Results Compared With Younger Patients: A Case-Control Study. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2019 8;35:2448-58.

[35] Ghisa C, Zaslav KR. Current state of off the shelf scaffolds and implants for meniscal replacement. Journal of Cartilage & Joint Preservation. 2022 3;2:100040.

[36] Martel-Pelletier J, Barr AJ, Cicuttini FM, Conaghan PG, Cooper C, Goldring MB, et al. Osteoarthritis. Nature Reviews Disease Primers. 2016;2(1):16072.

[37] Paradowski PT, Lohmander LS, Englund M. Osteoarthritis of the knee after meniscal resection: long term radiographic evaluation of disease progression. Osteoarthritis and Cartilage. 2016 5;24:794-800.

[38] Kohli S, Schwenck J, Barlow I. Failure rates and clinical outcomes of synthetic meniscal implants following partial meniscectomy: a systematic review. Knee surgery & related research. 2022 6;34:27.

[39] Wu KA, Therien AD, Kiwinda LV, Castillo CJ, Hendren S, Long JS, et al. Addressing meniscal deficiency part 2: An umbrella review of systematic reviews and meta-analyses on meniscal scaffold-based approaches. Journal of Experimental Orthopaedics. 2024 JUL;11(3).

[40] McCormack DJ, Puttock D, Godsiff SP. Medial compartment osteoarthritis of the knee: a review of surgical options. EFORT open reviews. 2021 2;6:113-7.

[41] Karsdal MA, Michaelis M, Ladel C, Siebuhr AS, Bihlet AR, Andersen JR, et al. Disease-modifying treatments for osteoarthritis (DMOADs) of the knee and hip: lessons learned from failures and opportunities for the future. Osteoarthritis and Cartilage. 2016;24(12):2013-21.

[42] Pabinger C, Lothaller H, Kobinia GS. Intra-articular injection of bone marrow aspirate concentrate (mesenchymal stem cells) in KL grade III and IV knee osteoarthritis: 4 year results of 37 knees. Scientific Reports. 2024;14(1):2665.

[43] Fusco G, Gambaro FM, Matteo BD, Kon E. Injections in the osteoarthritic knee: a review of current treatment options. EFORT open reviews. 2021 6;6:501-9.

[44] Smith AF, Eccles CJ, Bhimani SJ, Denehy KM, Bhimani RB, Smith LS, et al. Improved Patient Satisfaction following Robotic-Assisted Total Knee Arthroplasty. The Journal of Knee Surgery. 2021 6;34:730-8.

[45] DeFrance MJ, Scuderi GR. Are 20% of Patients Actually Dissatisfied Following Total Knee Arthroplasty? A Systematic Review of the Literature. The Journal of arthroplasty. 2023 3;38:594-9.

[46] Carballo CB, Nakagawa Y, Sekiya I, Rodeo SA. Basic Science of Articular Cartilage. Clinics in Sports Medicine. 2017 7;36:413-25.

[47] Segarra-Queralt M, Crump K, Pascuet-Fontanet A, Gantenbein B, Noailly J. The interplay between biochemical mediators and mechanotransduction in chondrocytes: Unravelling the differential responses in primary knee osteoarthritis. Physics of Life Reviews. 2024;48:205-21.

[48] Gilbert SJ, Blain EJ. Chapter 4 - Cartilage mechanobiology: How chondrocytes respond to mechanical load. In: Verbruggen SW, editor. Mechanobiology in Health and Disease. Academic Press; 2018. p. 99-126.

[49] Zhao Z, Li Y, Wang M, Zhao S, Zhao Z, Fang J. Mechanotransduction pathways in the regulation of cartilage chondrocyte homoeostasis. Journal of Cellular and Molecular Medicine. 2020;24(10):5408-19.

[50] Patwari P, Cook MN, DiMicco MA, Blake SM, James IE, Kumar S, et al. Proteoglycan degradation after injurious compression of bovine and human articular cartilage in vitro: Interaction with exogenous cytokines. Arthritis & Rheumatism. 2003;48(5):1292-301.

[51] Sanchez-Adams J, Leddy HA, McNulty AL, O'Conor CJ, Guilak F. The Mechanobiology of Articular Cartilage: Bearing the Burden of Osteoarthritis. Current Rheumatology Reports. 2014 10;16:451.

[52] D'Lima DD, Fregly BJ, Patil S, Steklov N, Colwell CW. Knee joint forces: prediction, measurement, and significance. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine. 2012 2;226:95-102.

Chapter 3

Research objectives

For this joint PhD in Medical Sciences and Physics, two main aspects were present, related to the two research domains. Advanced image processing techniques and extensive automation of time-consuming tasks such as image segmentation were essential to allow future clinical implementation and address the clinical challenges. Thus, in line with the clinical background of the previous chapters, the clinical objectives are described first, followed by a section on the required technology to address these clinical challenges.

3.1 Clinical objectives

3.1.1 3D knee morphology as risk factor for symptomatic knee degeneration in the meniscus deficient knee

Multiple risk factors contributing to the onset and progression of knee degeneration have already been described in current literature. While the genetic aspect certainly influence the tissue composition and quality, the focus in this work lies on the mechanical aspect. Known risk factors are mainly related to the load (distribution) applied on the knee joint: patient weight, coronal malalignment, activity patterns and traumatic events. This however does not explain the observed inter-individual variability in onset and progression rate to the fullest extent. A small subgroup of patients comes into clinical practice with a degenerative knee at a relatively young age with none of the previously described risk factors. Therefore, not only the applied load should be taken into account, but also the individual capacity to bear that load.

In the absence of other contributing risk factors, bony knee morphology can play a pronounced role in meniscus-deficient knees. Without a functional meniscus—whether due to a prior partial meniscectomy, traumatic injury, or degenerative lesion—the bone and cartilage are left as the primary structures to bear compressive and shear forces in the knee. Based on the discussion in Section 2.5: Consolidating mechanical risk factors in knee OA, we propose the hypothesis that symptomatic meniscus-deficient knees will be smaller, or at least exhibit a smaller affected knee compartment, than asymptomatic knees, assuming no additional mechanical risk factors, such as malalignment of the lower limb, are present.

3.1.2 Optimal patient selection for arthroscopic partial medial meniscectomy

Despite the shift in clinical practice toward preserving the meniscus as much as possible, there are still valid indications for an arthroscopic partial medial meniscectomy, as discussed earlier. The key to improving the success rate of this procedure may lie in refining the patient selection. A more precise, individualized approach, such as a prognostic model that estimates a patient's individual risk of developing the medial post-meniscectomy syndrome, could represent an important advancement. This model would allow clinicians to identify patients at high risk of failure and suggest alternative treatments where appropriate, thereby reducing the occurrence of poor outcomes.

The research described in this dissertation was conducted within the context of the MEFISTO project, which aimed to develop innovative solutions for meniscal replacement, including a bioactive resorbable meniscal scaffold and bioactive unloading prosthesis (artificial meniscus implant). A key contribution of this research was the development of a predictive algorithm based on the patient's morphotype, designed to predict the outcome of arthroscopic partial medial meniscectomy. The clinical relevance of this algorithm is significant, as it marks a crucial first step toward a data-driven, personalized treatment approach for managing meniscal injuries. By identifying patients at elevated risk of meniscectomy failure, the model could help guide clinicians toward meniscal substitution procedures or other alternatives, potentially preventing the onset of post-meniscectomy complications (see section 2.3.3: Prevention of the post-meniscectomy syndrome).

Given that meniscal substitution is typically reserved as a final option before joint replacement in symptomatic meniscus deficient patients (see section 2.3.2: Treatment options for post-meniscectomy syndrome), the predictive algorithm must demonstrate both high sensitivity and specificity. Sensitivity is crucial for identifying patients who are at risk of developing complications after meniscectomy, thus reducing the rate of failed primary surgeries. However, the scarcity in meniscus allograft donors and the relatively high failure rates of other meniscal substitution techniques necessitate caution. These procedures should only be considered for patients at substantial risk of meniscectomy failure, ensuring that the benefits of substitution outweigh the complication or failure risks. Therefore, while sensitivity is important for capturing all potential post-meniscectomy syndrome cases, the algorithm must also maintain high specificity to prevent overuse of meniscal substitution treatments.

The development of an accurate, morphotype-based predictive model for arthroscopic partial medial meniscectomy represents a fundamental innovation in personalized orthopedic care. By improving patient selection, this model has the potential to enhance the outcome of one of the most frequently performed procedures in orthopedics worldwide and minimize or delay the need for more invasive procedures such as knee arthroplasty surgery.

3.2 Technological framework

In order to achieve the previously mentioned clinical objectives, automation is essential to enable the autonomous processing of large imaging databases and allow fully automated outcome prediction (see 3.2.4: Automated prediction of treatment response versus failure). The first step to automate involves the creation of accurate 3D bone and cartilage models form the raw imaging data, a process called ”image segmentation”. Next, anatomical landmarks provide a simple and easy to visualize method to measure distinct distances and angles of interest. Furthermore, statistical shape analysis allow to automatically quantify the bone shape by means of just a confined number of modes of shape variation.

3.2.1 Automated image segmentation

Manual image segmentation is a highly time-consuming process, thus a massive efficiency gain will be achieved by automation of this step. Distinct approaches of neural network architectures will be implemented and evaluated in a clinically relevant way, by taking into account the anatomical location of eventual segmentation errors. Additionally, the required time for training and inference will be compared over the different network architectures. Following networks will be trained and evaluated: nnUNet, MedNeXt, SwinUNetR and SegMamba.

3.2.2 Automated anatomical landmarking

Similarly, automation of anatomical landmark annotation in turn allows for the automated calculation of landmark-based measures such as distances or angles. Extensive validation of both the manual and automated method for morphometric measurement analysis will be performed. Intra- and inter-observer variability will be compared against inter-method variability both for positioning the anatomical landmarks and obtaining the morphometric measurements. As the automated landmarking process is mainly based on the same registration process of the statistical shape analysis, this validations study will also ensure accurate point correspondences further used in the statistical shape analysis.

3.2.3 Automated statistical shape analysis

As an alternative to anatomical landmark-based measurements, statistical shape analysis allows to automatically quantify the bone shape by means of a confined number of modes of shape variation. It captures shape variations over the complete bone surface, thereby not limited to predefined landmarks, resulting in the detection of more challenging patterns of shape variation.

3.2.4 Automated prediction of treatment response versus failure

Finally, extracted bone shape features are integrated together with the clinical features of interest (e.g. patient sex, weight,...) into a predictive framework for response versus failure to treatment.

3.3 Outline of the thesis

In this opening part of the thesis, the clinical challenge was defined and situated. Elementary knee anatomy background was provided in chapter 1 to support the subsequent description of the pathophysiology related to meniscus and cartilage degeneration in chapter 2. The clinical objectives and technological framework of this PhD thesis are then introduced in the current chapter 3. The second main part of this thesis is dedicated to the development and implementation of a technological framework to address the clinical challenge. The first two chapters of this part explain the need for a more efficient workflow and the evolution from a manual to a fully automated approach to create accurate 3D bone models from raw imaging. Next, four different neural network architectures to do this are discussed in chapter 6. These can be employed to identify the anatomy of interest in the raw images. Consequently, the automated process of anatomical landmark annotation based on these 3D bone models is described in chapter 7, including an extensive validation study. As the bone shapes in this study will be quantified not only by landmark-based measures, but also by a more general statistical shape analysis, the technical background in chapter 8 aims to provide a deeper understanding of this method. Both methods rely on the same surface registrations step as detailed in section 8.3. Finally, in order to link a combination of clinical (patient age, sex, weight,...) and morphological features to the treatment outcome, a predictive model will be developed using a machine learning approach. Some elementary concepts of machine learning are therefore described in chapter 9. The third main part then describes the clinical applications of the previously illustrated technological framework. Chapter 10 illustrates how to choose the best neural network architecture for the segmentation task at hand (in this case: bone and cartilage from knee MRI). The next chapter shortly describes two distinct applications of the surface registration process described in section 8.3: one involves an automated method to measure the femoral version and the other includes the estimation of the meniscus geometry based on the knee bone geometry. The next two chapters 12 and 13 explore studies demonstrating the potential role of knee morphology in the early onset and rapid deterioration of knee degeneration, in the absence of other known risk factors. Chapter 14 then highlights the observed morphological differences between the knees with APMM response to treatment and the knees with MPMS. Next, the development of a predictive model for MPMS in APMM patients is discussed in chapter 15. The final part then includes a concluding chapter, as well as a final chapter describing the future perspectives.

Part II

Methodological framework

Technical background on the image acquisition, processing and analysis for developing of a morphotype-based prognostic model for arthroscopic partial medial meniscectomy.

Chapter 4

Image modalities

All patient-specific 3D models are derived from appropriate imaging. In this chapter, some of the most frequently used image modalities in orthopaedics are explored, such as computed tomography (CT) and magnetic resonance imaging (MRI). Furthermore, some imaging key characteristics that have an influence on the subsequent segmentation process are discussed. Although ultrasound (US) imaging presents advantages such as low cost and high accessibility, it is not further discussed here. For anatomical morphology analysis, US imaging introduces potential operator-dependent variability, which may affect consistency. Furthermore, the retrospective interpretation of US images is challenging without live patient interaction or precise information about the spatial positioning of the US probe. Therefore, it is less suited for the purposes of this study.

4.1 Computed Tomography

Computed Tomography (CT) provides detailed cross-sectional images of the human body using X-ray technology.

4.1.1 Basic principles of CT imaging

CT imaging operates on the fundamental principle of differential X-ray attenuation by various tissues. As X-rays pass through the body, they are absorbed at different rates depending on the tissue material properties and thickness (Figure 4.1). In a very simplified form (for homogeneous materials, and neglecting any scattering), the X-ray intensity measured by the detector can be calculated by Beer-Lambert's law (Eq. 4.1). Thus, the intensity of the X-rays on the detector is mainly dependent on the tissue characteristics and thickness, next to the initial X-ray intensity (I_0) [1].

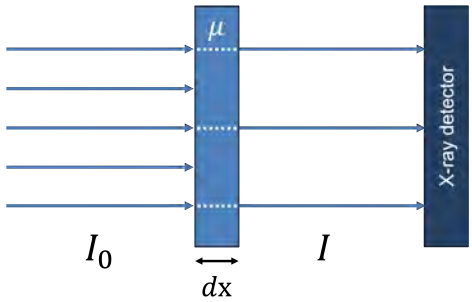


Figure 4.1: Simplified illustration of X-ray attenuation. The initial beam of X-rays has an intensity I_0 and the intensity I decreases exponentially with the distance dx travelled inside the tissue. The material property μ is responsible for the differential X-ray absorption and intensity at the X-ray detector for distinct tissues.

$$I = I_0 e^{-\mu dx} \tag{4.1}$$

CT scanners produce images through the rotation of an X-ray source and detector around the patient. The detector measures the attenuated X-rays, and sophisticated algorithms reconstruct cross-sectional images from this raw scan data sinogram (Figure 4.2). The acquisition time of CT scans is considered fast and patient movement artifacts are relatively rare. Resulting images exhibit varying shades of gray, correlating to tissue X-ray absorption (and densities), facilitating the identification of anatomical structures and abnormalities. The shades of gray in the image represent Hounsfields Units (HU): the relative tissue radiodensity at a certain position in space. Therefore, HU is a quantitative measure, which scales as follows: 0 HU represents radiodensity of distilled water and -1000 HU represents the radiodensity of air at 0°C and 10^5 Pa.

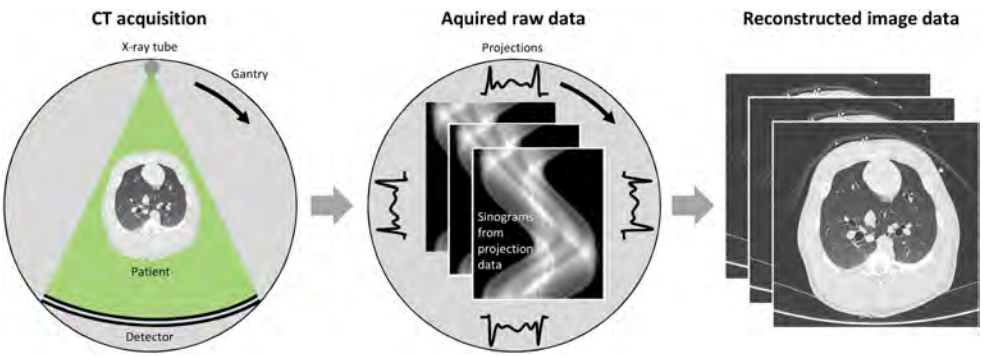


Figure 4.2: CT raw data acquisition and image reconstruction. [2].

4.1.2 Challenges in CT imaging

Despite its benefits, CT imaging still poses some challenges:

- Radiation exposure: Repeated exposure to ionizing radiation raises concerns about potential long-term risks, such as cancer, reduced fertility or genetic changes.

- Image artifacts: Artifacts from patient movement or metallic implants (resulting in X-ray scattering) can compromise image quality and interpretation.
- Less effective for soft tissue visualization, where the distinct tissues have similar radiodensities.

4.1.3 Applications of CT in orthopaedics

CT imaging serves multiple clinical purposes in orthopaedics, including but not limited to:

- Diagnosis of small bone fractures, which are difficult to detect on plain radiographs.
- Diagnosis and 3D surgical planning of malunion fractures (improper alignment of the bone parts during healing).
- 3D lower limb alignment analysis in the three planes: coronal, sagittal and axial.
- Patient-specific surgical planning (Figure 4.3) and 3D printing of patient-specific instruments (PSI)

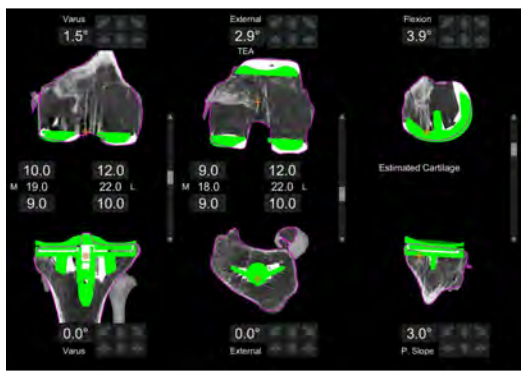


Figure 4.3: CT-guided planning of a complex total knee prosthesis revision case. Optimal size selection and positioning can be defined carefully pre-operatively and even further finetuned during surgery [3].

4.1.4 Conclusion

In conclusion, CT scans are the optimal choice for bone imaging or imaging that needs to be fast (emergencies, prevention of motion artifacts) and at an adequate resolution at the same time, when a limited radiation exposure is acceptable.

4.2 MRI

Magnetic Resonance Imaging (MRI) is a totally different technique, which produces also detailed anatomical images in three dimensions. Based on the principles of nuclear magnetic resonance (NMR), MRI exploits the magnetic properties of certain atomic nuclei (usually hydrogen) to visualize soft tissues in biomedical imaging.

4.2.1 Basic principles of MRI

At its core, MRI relies on the behavior of hydrogen nuclei, which are abundant in the human body (especially in soft tissues like muscles, ligaments and even cartilage) due to the high water and fat content. These hydrogen nuclei possess a property called spin, making them act like tiny magnets. When a patient is placed inside an MRI scanner, a powerful external magnetic field (B_0), typically 1.5 (most clinical settings) or 3.0 Tesla (often in university hospitals or research facilities), aligns the spins of these hydrogen nuclei to the magnetic field.

However, simply aligning the magnetic spins of the hydrogen nuclei is not sufficient for imaging. The key to MRI is perturbing this alignment with a radiofrequency (RF) pulse. This RF pulse is delivered at a specific frequency, known as the Larmor frequency f_0, which is determined by the strength of the magnetic field B_0 and the gyromagnetic ratio γ for the atom of interest (in our case hydrogen).

When this specific RF pulse is applied, it tips the aligned hydrogen nuclei out of their equilibrium position. Once the RF pulse is turned off, the hydrogen nuclei begin to relax back to their original alignment with the magnetic field. As they do so, they emit energy in the form of electromagnetic radio waves. These are then detected by the different coils (e.g. a multi-channel knee coil captures these radio waves in different orientations while scanning the knee) within the MRI scanner. The relaxation process occurs in two main ways (Figure 4.4 parts b and c): longitudinal relaxation (T1) and transverse relaxation (T2). T1 relaxation refers to the time it takes for the nuclei to realign with the magnetic field, while T2 relaxation refers to the time it takes for the nuclei to lose phase coherence among themselves.

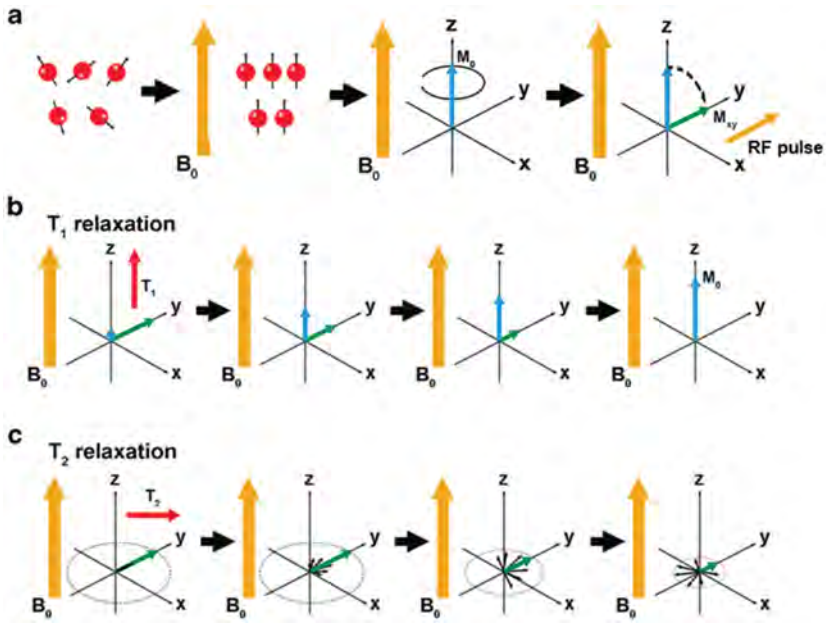


Figure 4.4: Schematic overview of the basic MRI principles. **a)** An external magnetic field B_0 (orange arrow) is applied (z-direction), and induces a net magnetization vector (M_0, blue arrow) aligned with B_0. When an orthogonal RF pulse is applied, M_0 tilts 90° into the transverse (xy) plane (M_{xy}, green arrow). The magnetization returns to equilibrium through two processes: T1 and T2 relaxation. **b) T1 relaxation:** T1 is a measure of the time it takes for the initial longitudinal magnetic moment (M_0) to recover. **c) T2 relaxation:** T2 measures the loss of the transverse magnetic moment (M_{xy}) due to dephasing [4].

The signals received by the coils are then processed to create images. In order to locate the signal origin, spatial encoding is performed by applying different magnetic field gradients in perpendicular directions, finally resulting in a raw k-space image. Next, this is converted into a spatial domain image by applying inverse Fourier transformation. In 2D MRI protocols, only signals from one specific slice at a time are captured. Therefore, if a high number of slices is required (e.g. large area or fine slices), the scan time increases linearly. Additionally, the size of the raw k-space is equivalent to the final in-plane image resolution. A larger k-space to be sampled might have more phase-encoding steps as a consequence, resulting in a longer scan time. Hence, the maximal achievable in-plane image resolution is related to the maximal feasible scan time (higher cost and risk for motion artifacts).

Recently, different acceleration algorithms have been developed to capture high-resolution (near-)isotropic images within a reasonable scan time, including but not limited to 3D-fast spin echo acquisition (SPACE/CUBE/VISTA depending on the MRI manufacturer), parallel imaging techniques (e.g. SENSE, GRAPPA, CAIPIRINHA) and also deep-learning techniques (e.g. Deep Resolve on Siemens MRI machines). While originating from research environments, these 3D MRI techniques were proven to perform interchangeably with standard 2D MRI protocols for diagnosing knee abnormalities in orthopaedic clinical practice [5]. As a result, they now start to find entrance into routine clinical use.

4.2.2 Thick-Slice 2D MRI / Thin-Slice 3D MRI

MRI sequences can be categorized into thick-slice 2D and thin-slice 3D acquisition protocols, each having distinct characteristics and applications.

2D MRI

- Acquires images slice by slice, each with a thickness of ca. 3 to 5 mm, commonly used in clinical practice (Fig. 4.5 top row).
- **Advantages:**
  Faster acquisition time and hence lower cost, useful for initial surveys and scanning of large areas.
- **Limitations:**
  By definition they provide a lower resolution in the slice-selection direction. As a result, they are more prone to partial volume artifacts. These typically occur when the sampled spatial location includes multiple tissues characterized by distinct MRI signals. Finally, the low resolution in one direction makes them less accurate for detailed anatomical studies. Hence, multiple sequences in perpendicular acquisition planes are desirable to create accurate and smooth 3D models of the anatomy of interest, in order to avoid coarse stair-step effects, caused the discrete nature of the voxel grid in the slice direction.

3D MRI

- Acquires slabs of multiple slices at once as volumetric data, allowing for thin-slice reconstructions in any plane (Fig. 4.5 bottom row), currently mainly used in research and neuro-imaging.
- **Advantages:**
  Higher spatial resolution, better anatomical detail, fewer artifacts. Additionally these sequences are superior for multiplanar reconstructions and complex anatomical areas. Only one scan sequence is needed to create accurate 3D models of the anatomy of interest. Multiple sequences with different echo and repetition times (TE and TR) might be useful to obtain multiple distinct tissue contrasts for easier pathology diagnosis or anatomy delineation.
- **Limitations:**
  Stricter hardware requirements, longer acquisition times, higher sensitivity to patient motion due to the longer scan time, and increased computational requirements for data processing.

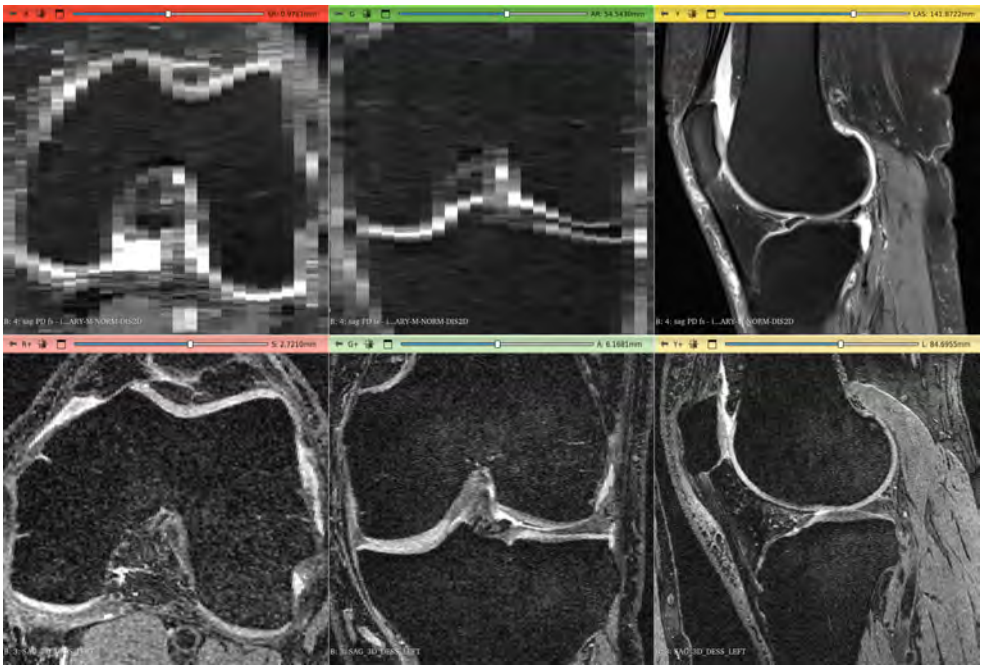


Figure 4.5: Top: example of a standard clinical 2D MRI sequence with a high in-plane sagittal (right) resolution, but a rather course resolution in the axial (left) and coronal (middle) cross sections. **Bottom**: example of a near-isotropic 3D MRI sequence, voxel dimensions are below 1mm, resulting in adequate cross-sections in the three main anatomical planes (left, middle and right). Note that not only the scan resolution differs between these two sequences, also the image contrast is different due to different acquisition parameters (proton density weighted versus dual echo steady state).

4.2.3 Challenges in MRI Imaging

Despite its benefits, MRI still poses some challenges:

- Artifacts: distortions or errors in MRI images that can obscure or mimic pathology. Common artifacts include:
  - Motion artifacts: Caused by patient movement during the scan, leading to blurred images.
  - Magnetic susceptibility artifacts: Result from variations in the magnetic properties of different tissues or the presence of metal objects, disturbing the homogeneous magnetic field.
- Cost and Accessibility: MRI is costly, both in terms of initial setup, scan time (operators and staff) and ongoing maintenance. This can limit accessibility, particularly in low-resource settings.
- Safety Concerns: The strong magnetic fields used in MRI can pose safety risks, especially for patients with metal implants or other magnetic-sensitive devices. Thorough pre-scan screening is required to ensure patient safety.

- The trade-offs between scan time (typically 15-30 minutes per MRI of the knee), scan resolution and signal-to-noise ratio (Figure 4.6). While higher resolution and signal to noise ratio (SNR) are desirable for accurate diagnosis, they often necessitate longer scan durations, leading to increased costs and higher chances of motion artifacts due to patient movement during the scan. Moreover, any adjustment to scan parameters that affects either scan resolution or SNR will inevitably influence at least one of the other two factors. This intricate interplay highlights the challenge of optimizing image quality while considering the practical constraints of scan time and cost.

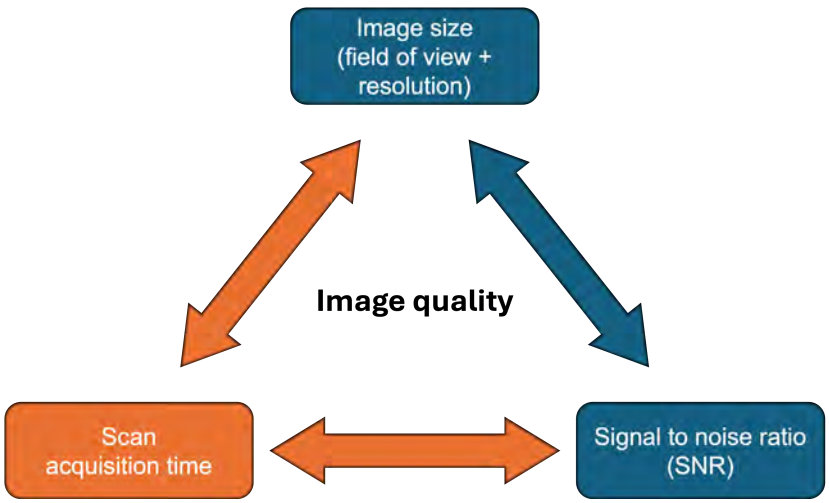


Figure 4.6: Trade-offs to consider during MRI scan protocol design. Image size is determined by the field of view and the scan resolution (voxel size). A high signal to noise ratio enhances differential tissue contrast, facilitating the delineation of distinct anatomical structures. Finally, minimization of the scan acquisition time is important to reduce costs and risk for motion artifacts.

4.2.4 Applications of MRI in Orthopaedics

While conventional radiographs or CT are the best-fit imaging modality for diagnosing bone fractures, MRI is better-fit for soft tissue pathologies. Moreover, MRI does not utilize ionizing radiation, being a safer options for multiple repetitive scans and annual screenings compared to CT scans.

MRI is particularly valuable for knee specialists because it provides high-resolution images of soft tissues, such as muscles, ligaments, tendons, cartilage and the menisci. The versatility of MRI sequences, such as T1-weighted, T2-weighted, and proton density (PD) sequences, allow for detailed evaluation of various conditions. For instance, T1-weighted images are useful for visualizing the anatomy and fatty tissues, while T2-weighted images highlight areas with high water content, such as edema or inflammation, making it easier to identify injuries and pathological changes.

MRI has multiple applications in orthopaedics, including but not limited to:

- Diagnosis of Soft Tissue Injuries: MRI is particularly effective for imaging soft tissues, making it the preferred method for diagnosing muscle, ligament, and tendon injuries. For example, MRI can provide detailed images of rotator cuff tears in the shoulder, meniscal tears in the knee, and other soft tissue pathologies.
- Evaluation of Bone Pathologies: MRI can reveal some bone pathologies that are not visible on radiographs or CT scans. Conditions such as bone marrow edema, osteonecrosis, and stress fractures are often easier to assess with MRI.
- Preoperative Planning and Postoperative Assessment: MRI plays a crucial role in preoperative planning by providing surgeons with detailed anatomical maps (e.g. for mini-implants in the treatment of large focal cartilage defects). Postoperatively, MRI is used to assess the success of surgical interventions and monitor the healing process.
- Monitoring of Chronic Conditions: MRI is useful for monitoring chronic orthopedic conditions such as rheumatoid arthritis and osteoarthritis. It can assess the extent of joint damage, inflammation, and other pathological changes over time.

4.2.5 Conclusion

MRI is a powerful and versatile imaging modality that has revolutionized the field of medical imaging. Despite its challenges, including scan artifacts, scan time and cost, its unparalleled ability to visualize soft tissues and some specific bone pathologies makes it an indispensable tool in orthopaedic clinical practice. As technology advances, continued improvements in MRI techniques and accessibility will enhance its clinical utility and expand its applications further.

4.3 Image modalities summary

	CT	MRI
Principle	Differential X-ray attenuation	Magnetic resonance of hydrogen nuclei
Challenges	Radiation Scatter artifacts (metal) Soft tissue image contrast	Cost Magnetic susceptibility artifacts (metal: field inhomogenities) Motion artifacts Trade-off scan time/resolution/SNR Safety (MRI compatibility)
Applications	Bone imaging 3D anatomy modelling	Soft tissue imaging incl. bone marrow (fatty tissue, H-atoms) 3D anatomy modelling (3D scan protocols)

Bibliography

[1] Baur M, Uhlmann N, Pöschel T, Schröter M. Correction of beam hardening in X-ray radiograms. Review of Scientific Instruments. 2019 2;90.

[2] Stiller W. Basics of iterative reconstruction methods in computed tomography: A vendor-independent overview. European Journal of Radiology. 2018 12;109:147-54.

[3] Ngim HLJ, Bavel DV, Steiger RD, Tang AWW. Robotic-assisted revision total knee arthroplasty: a novel surgical technique. Arthroplasty. 2023 1;5:5.

[4] Mastrogiacomo S, Dou W, Jansen JA, Walboomers XF. Magnetic Resonance Imaging of Hard Tissues and Hard Tissue Engineered Bio-substitutes. Molecular Imaging and Biology. 2019 12;21:1003-19.

[5] Dyck PV, Smekens C, Roelant E, Vyvere TV, Snoeckx A, Smet ED. 3D CAIPIRINHA SPACE versus standard 2D TSE for routine knee MRI: a large-scale interchangeability study. European Radiology. 2022 3;32:6456-67.

Chapter 5

Image segmentation and 3D meshing

In this chapter, an introduction will be given how 3D anatomy models are created from biomedical imaging by means of a manual method. Multiple tools exist to assist the human operator to achieve a faster and more efficient workflow. As these tools and software packages still require human operator input, they are labeled as manual segmentation tools. For matters of completeness, the 3D meshing of the segmented image volume as final step will only be shortly described (5.2: 3D meshing) as this commonly used algorithm is not the focus of this work.

5.1 Manual image segmentation

5.1.1 Introduction

Image segmentation is a computer vision task that involves dividing an image or a scan volume into distinct regions, or segments, based on specific criteria. In medical imaging, this process is performed for identifying and delineating anatomical structures, lesions, or abnormalities within medical images. By accurately segmenting these regions, physicians can gain valuable insights into disease diagnosis, treatment planning, or monitoring disease progression.

Image segmentation algorithms employ various techniques, including thresholding, edge detection, and machine learning approaches. Thresholding involves dividing an image into regions based on intensity values, while edge detection aims to identify boundaries between distinct anatomical regions. Machine learning algorithms, particularly deep learning models, have shown remarkable success in image segmentation by learning complex patterns and features within medical images (see Chapter 6. Automated image segmentation: neural networks). They provide a huge efficiency gain in this highly time-consuming task when performed manually [1]. In the context of this PhD research project, 240 clinical MRI scans of the knee were manually segmented, which took over a year to assemble the complete 3D shape database of post-meniscectomized distal femora and proximal tibiae.

Because biomedical image segmentation per definition only encompasses the delineation of anatomical structures in the image or scan domain, thus classifying each

voxel as belonging to the structure or not, an additional step is needed to create a 3D mesh suitable for 3D shape analysis (see section 5.2. 3D meshing). A 3D mesh is a digital representation of three-dimensional object, as a network of interconnected points (or vertices, these terms will be used interchangeably troughout this work) defining the shape and surface of the object. Depending of the type of mesh it can be used for visualization purposes, as input for 3D printing (typically triangulated surface meshes) or in-silico biomechanical simulations such as in finite element analysis (typically volumetric hexahedral meshes).

5.1.2 Software for manual image segmentation

While the basic nature of this task (classifiying voxels as belonging to the anatomy of interest or not) does not necessitates advanced computations or algorithms, all available (both commericial and open-source) software packages offer some tools to accelerate this highly time-consuming task [2]. Some of the most useful tools include:

- **Image registration**: this technique aligns multiple images or sequences from different sources or time points, ensuring consistency and accuracy in annotations across the subject by moving all subjects or images into the same reference coordinate system. Using the same reference coordinate system is crucial when information or image features from different images or scan sequences are combined.
- **Thresholding**: by setting a threshold value, voxels with intensities above or below the threshold can be automatically classified as belonging to or not belonging to the anatomy of interest. This can be a quick and effective method for simple structures and works especially well in low-noise CT images.
- **Region growing**: this algorithm starts from a seed voxel and iteratively expands the region based on similarity criteria, such as intensity or texture. It can be useful for segmenting connected regions with a comparable appearance within an image.
- **Slice interpolation**: this technique creates intermediate slice annotations between annotated slices. It is extremely helpful to reduce the number of slices that need fully manual segmentation.
- **Particle removal**: for images with noise or artifacts that appear as small, isolated objects, particle removal algorithms can help clean up the image and improve annotation accuracy.
- **Morphological operations**: such as closing (filling holes) or opening (removing small objects), can be used to refine the segmentation results and improve the quality of the annotated regions.
- **Boolean operations**: such as union, intersection or subtraction allow to perform logical operations with segmentation masks from multiple anatomies or objects.

5.1.3 Challenges and pitfalls

Segmenting MRI scans from multiple thick-slice sequences is an iterative process: changes made while inspecting one sequence might introduce new out-of-plane small errors. Manual segmentation of thick-slice 2D MRI knee protocols imposes several challenges due to the inherent limitations of this imaging modality. These challenges can significantly impact the accuracy, efficiency, and reproducibility of segmentation tasks.

2D MRI Segmentation: Time Consumption and Labor Intensity

The process of manually tracing anatomical structures in thick-slice images is time-consuming and labor-intensive, particularly for relatively thin knee structures such as cartilage, menisci and ligaments. Moreover, to ensure an accurate image segmentation also in the out-of-plane direction, multiple MRI sequences (with perpendicular acquisition planes) are necessary since interpolation over the interslice distance of ca. 3,5mm potentially leads to inaccuracies for thin anatomical structures (e.g. cartilage)[3]. When using multiple MRI sequences, it is crucial the patient did not move between the acquisition of the sequences. An additional image registration step can ensure all sequences to be aligned within the same reference coordinate system, but adds onto the already high labor-intensity and time-consumption. Typically, the manual segmentation of distal femur (bone + cartilage in a single mask) and proximal tibia (bone + cartilage in a single mask) from a clinical thick-slice 2D MRI scan takes between 2,5 and 5 hours, depending on the total image quality.

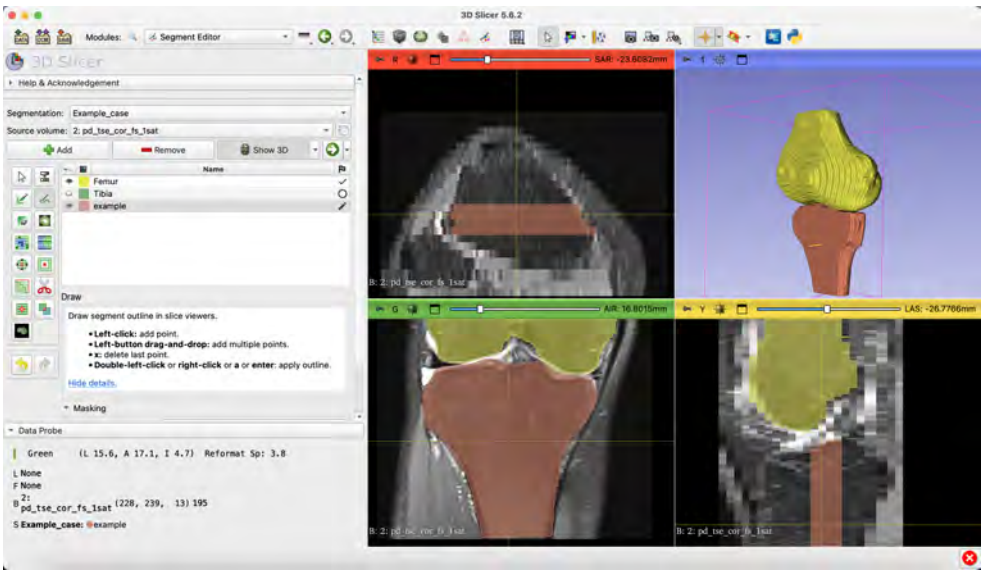


Figure 5.1: Screenshot from 3D Slicer [4] during the first step of image segmentation from a thick-slice scan protocol. A clear staircase effect is visible on the preliminary 3D visualization in the upper right corner, due to the large slice thickness.

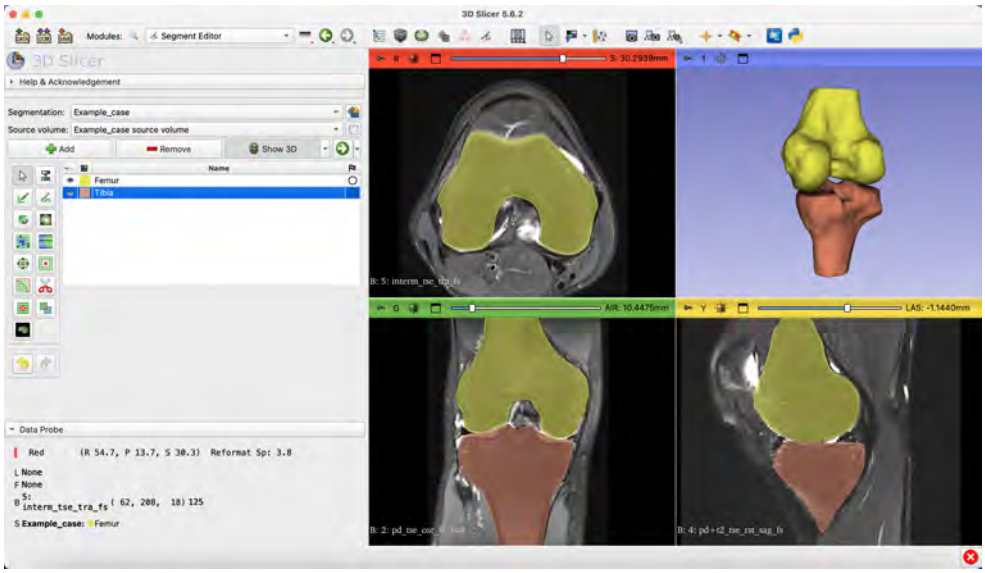


Figure 5.2: Screenshot from 3D Slicer [4] at the end of the image segmentation process. By iteratively refining the segmentation over all available scan sequences (perpendicular acquisition planes), a smooth and accurate final 3D model could be achieved.

Image Quality Limitations

- **Partial Volume Effects:** Thick slices (and thus large image voxels) are more prone to partial volume effects, where a single voxel contains contributions from multiple tissues. This makes it challenging to accurately identify tissue boundaries and can potentially introduce errors in segmentation. For example, in a thick slice of the knee joint, a single voxel might contain contributions from both cartilage and bone or synovial fluid, making it challenging to distinguish between these tissues.
- **Low Contrast**: In some cases or scan protocols, images may have low contrast between different tissues, making it challenging to distinguish them and increasing the risk of segmentation errors. This can be particularly problematic in regions with subtle anatomical variations or in the presence of imaging artifacts. A typical example of this was sometimes encountered at the region of the tuberositas tibiae, at the border between tibial cortex and patellar tendon.
- **Artifacts**: Imaging artifacts, such as motion artifacts, magnetic susceptibility artifacts (e.g. due to the presence of metal), or noise, can further complicate the segmentation process, particularly in regions with low signal-to-noise ratio (e.g. scan volume borders). These artifacts can obscure tissue boundaries and make it difficult to accurately delineate anatomical structures.

Segmentation Accuracy and Reproducibility

- **Suboptimal Segmentation**: Thick slices may not capture the fine anatomical details required for accurate segmentation, leading to suboptimal results. This can

be particularly problematic for studies that require precise measurements of thin anatomical structures, such as cartilage thickness or meniscus morphology [5, 6].

- **Inter-Observer and Intra-Observer Variability**: variations in user interpretation and performance can impact both inter-observer and intra-observer variability, reducing the reliability of segmentation results. This can make it difficult to compare results across different studies or to track changes in a patient's condition over time.

To address these challenges, researchers and clinicians are increasingly exploring automated segmentation methods using machine learning and deep learning techniques. These methods have the potential to significantly reduce the time required for segmentation while improving accuracy, consistency, and reproducibility. By automating the segmentation process, we can streamline clinical workflows, enhance the reliability of quantitative measurements, and facilitate more efficient and accurate diagnosis and treatment of knee joint pathologies.

5.2 3D meshing

The marching cubes algorithm is one of the most popular techniques for converting a 3D binary segmentation mask into a triangulated mesh representation. It operates by examining each voxel in the mask (Figure 5.3) and determining the configuration of its neighbors. Based on this configuration, a table of pre-computed triangle vertices is used to generate a set of triangles that approximate the surface of the segmented object. By iterating through all voxels in the mask, the algorithm constructs a complete 3D mesh representing the segmented region (Figure 5.4). The marching cubes algorithm is an efficient and widely used method in medical imaging and other fields where 3D visualization and 3D surface analysis of segmented objects are required.

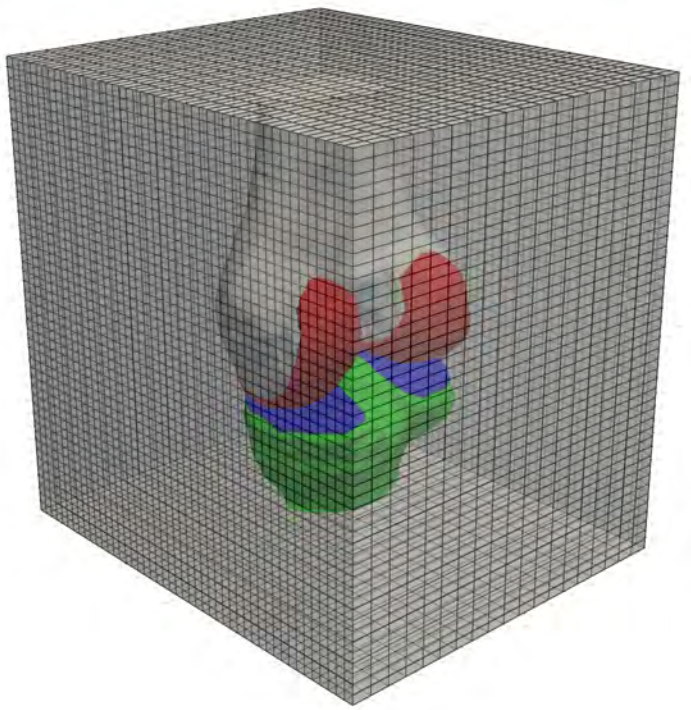


Figure 5.3: Overlay of the voxel grid of the scan volume containing the 3D anatomy of interest. The marching cubes algorithm runs over the complete image volume voxel by voxel to create a surface mesh of the anatomy of interest.

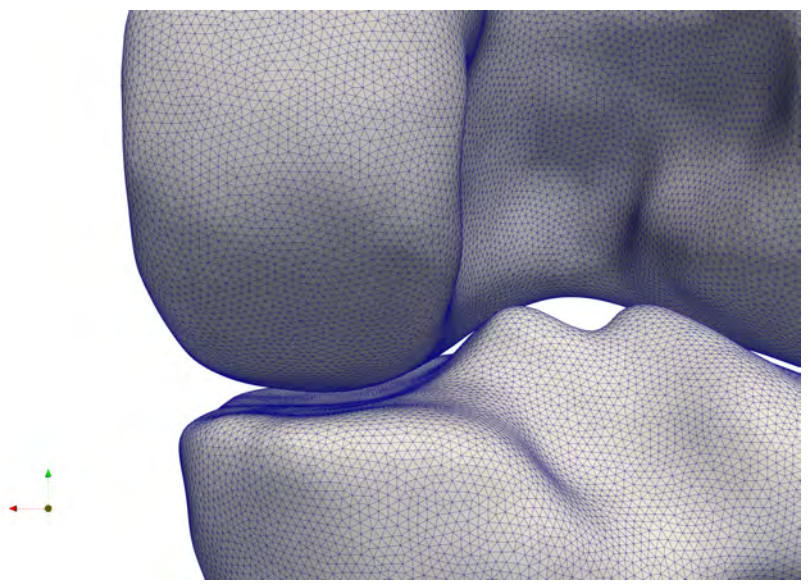


Figure 5.4: Detail of a triangular 3D mesh structure of distal femur (upper mesh) and proximal tibia (lower mesh) from a posterior point of view.

Bibliography

[1] Kessler DA, MacKay JW, McDonnell SM, Janiczek RL, Graves MJ, Kaggie JD, et al. Segmentation of knee MRI data with convolutional neural networks for semi-automated three-dimensional surface-based analysis of cartilage morphology and composition. Osteoarthritis Imaging. 2022 6;2:100010.

[2] Slicer Community. Image Segmentation; 2022. `https://slicer.readthedocs.io/en/latest/user_guide/image_segmentation.html` [Accessed: 2 Oct 2024].

[3] Beeler S, Vlachopoulos L, Jud L, Sutter R, Fürnstahl P, Fucentese SF. Contralateral MRI scan can be used reliably for three-dimensional meniscus sizing — Retrospective analysis of 160 healthy menisci. The Knee. 2019 10;26:954-61.

[4] 3D Slicer contributors, Brigham and Women's Hospital (BWH). 3D Slicer; 2022. `https://www.slicer.org` [Accessed: 2 Oct 2024].

[5] Okazaki Y, Furumatsu T, Yamaguchi T, Kodama Y, Kamatsuki Y, Masuda S, et al. Medial meniscus posterior root tear causes swelling of the medial meniscus and expansion of the extruded meniscus: a comparative analysis between 2D and 3D MRI. Knee Surgery, Sports Traumatology, Arthroscopy. 2020 11;28:3405-15.

[6] Kayfan S, Hlis R, Pezeshk P, Shah J, Poh F, McCrum C, et al. Three-dimensional and 3-Tesla MRI morphometry of knee meniscus in normal and pathologic state. Clinical Anatomy). 2021 1;34:143-53.

Chapter 6

Automated image segmentation: neural networks

In this chapter, some elementary concepts of neural networks for computer vision applications are explained. Furthermore, some design rationales for four recent state-of-the-art neural network architectures are discussed. These neural networks are implemented for automated knee MRI segmentation. Resulting performance and accuracy are compared over these four neural network architectures in chapter 7: Statistical parametric mapping for segmentation evaluation.

6.1 Introduction

The substantial time required for manual segmentation of knee MRI scans has prompted a search for innovative solutions to address this issue, while maintaining at least an equal level of accuracy and practical usability in clinics. Recently, significant advancements in computer vision have been made, driven by the availability of large datasets, open-source AI libraries like PyTorch [1] and TensorFlow [2], and the continuous increase in computing power. Today's neural networks, when provided with suitable training data, are already well-equipped to handle highly repetitive tasks such as image segmentation. The implementation of these innovative approaches for the initial step in morphological analysis (=image segmentation), leads to a substantial efficiency boost. This, in turn, by removing the first bottleneck, allows to scale the process of anatomical morphology analysis to nearly unlimited dataset sizes. In this chapter, we introduce the basic concept of neural networks and illustrate the evolution over time in neural network architectures for image segmentation by some examples. In a following chapter (Chapter 7. Statistical parametric mapping for segmentation evaluation), some of these examples were implemented for segmentation of knee MRI scans, and evaluated in a clinically relevant way by taking into account the spatial (anatomical) location of eventual segmentation errors.

6.2 Neural networks

Neural networks are computational models drawing inspiration from the human brain's architecture. Fundamentally, they consist of interconnected nodes, or neurons, arranged in layers. The very first related implementation was the perceptron algorithm, which is essentially a neural network consisting of 1 node (and thus also only one layer). The analogy between a biological neuron and the perceptron algorithm is illustrated in Figure 6.1.

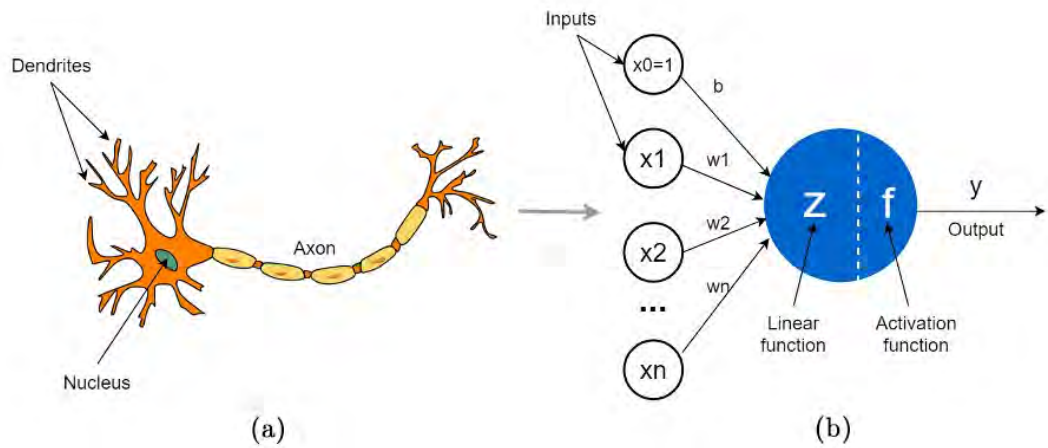


Figure 6.1: (a): Drawing of a biological neuron. It receives input from other neurons trough its dendrites, the output signal is produced in the nucleus and transmitted trough the axon. **(b)**: Schematic representation of a perceptron node. This perceptron node receives inputs from other perceptron nodes (denoted as x0, x1, ..., xn) and applies weights (b, w1, ..., wn) to these inputs. The weighted inputs are processed through a linear function followed by a (non-linear) activation function to produce the output of the node. In this analogy, the inputs to the perceptron are equivalent to the dendrites of the biological neuron, the linear and activation functions of the perceptron mirror the processing function of the nucleus, and the output of the perceptron corresponds to the function of the axon in the neuron. Figure adapted from [3].

6.2.1 Deep learning

A neural network with many (hidden) layers, is also called a deep neural network (Figure 6.2). The concept of deep learning refers to these specific machine learning architectures. Data is fed into the network, traversing these layers and undergoing mathematical transformations. The strength of connections between neurons is quantified by numerical values known as weights. Through a process analogous to human learning, these weights are adjusted iteratively to optimize the network's ability to recognize patterns within the data. In essence, a neural network is a sophisticated mathematical function capable of approximating intricate relationships between input and output variables. This capacity to discover complex and non-linear patterns in

voluminous datasets makes neural networks a powerful tool for many applications, including image and speech recognition, natural language processing, and, increasingly, diagnostic and predictive models in medicine.

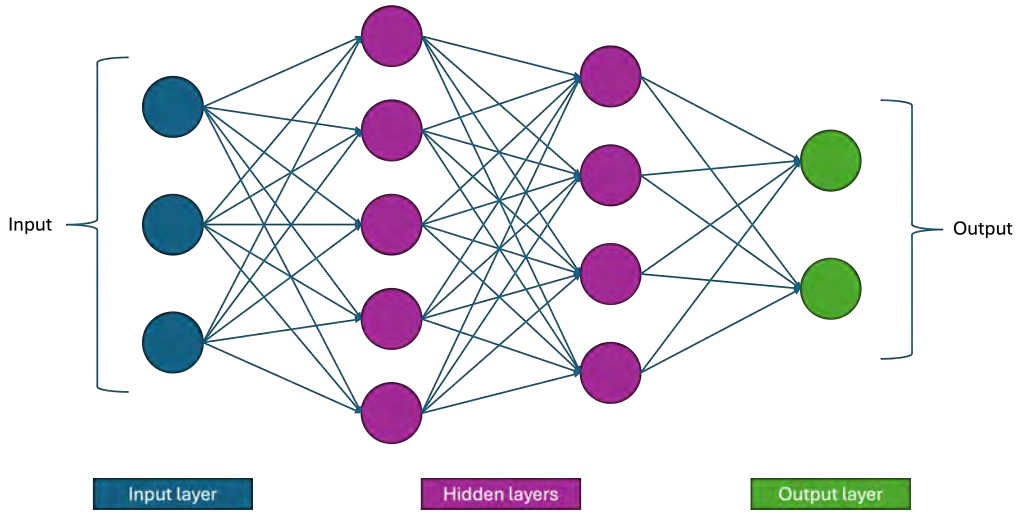


Figure 6.2: Illustration of a neural network with two hidden layers. This is one of the simplest neural network architectures: a multilayer perceptron (MLP) [4].

While the human brain employs biological structures for communication between neurons, artificial neural networks utilize mathematical operations. The process of learning, or training, in a neural network involves exposing it to large quantities of labeled data and iteratively refining the weights to minimize prediction errors. This adaptation is achieved through algorithms that calculate the error between the network's output and the desired outcome (loss function), and subsequently adjust the weights accordingly during a process called backpropagation.

6.3 U-Net

The U-Net architecture, as introduced by Ronneberger et al. in their 2015 paper "U-Net: Convolutional Networks for Biomedical Image Segmentation" [5], has become a fundament in the field of automated image segmentation. The architecture is named "U-Net" due to its distinctive U-shaped structure, which facilitates the capture of both local and global features or image characteristics necessary for an accurate delineation of the objects of interest.

The U-Net architecture consists of two main parts: a contracting path (encoder) and an expansive path (decoder) (Figure 6.3). The contracting path follows the typical architecture of a convolutional neural network (CNN). It comprises a series of convolutional layers with small 3x3 kernels, followed by activations (ReLu) and 2x2 max-pooling operations with a stride of 2. These convolutional operations act like distinct image filters to extract multiple image features and progressively reduce the

spatial dimensions of the input image while increasing the number of feature channels. Each feature channel then captures some specific image characteristics as calculated by the corresponding convolutional filter. This process allows the network to capture and learn increasingly abstract and complex features of the input data. In conclusion, the encoder serves to extract essential contextual information from the input images, which is critical for accurate segmentation. Variants on the default parameters (kernel size, stride, type of activation function...) have been investigated (within the limitations posed by the currently available hardware) and their success rates depend on the specific task at hand.

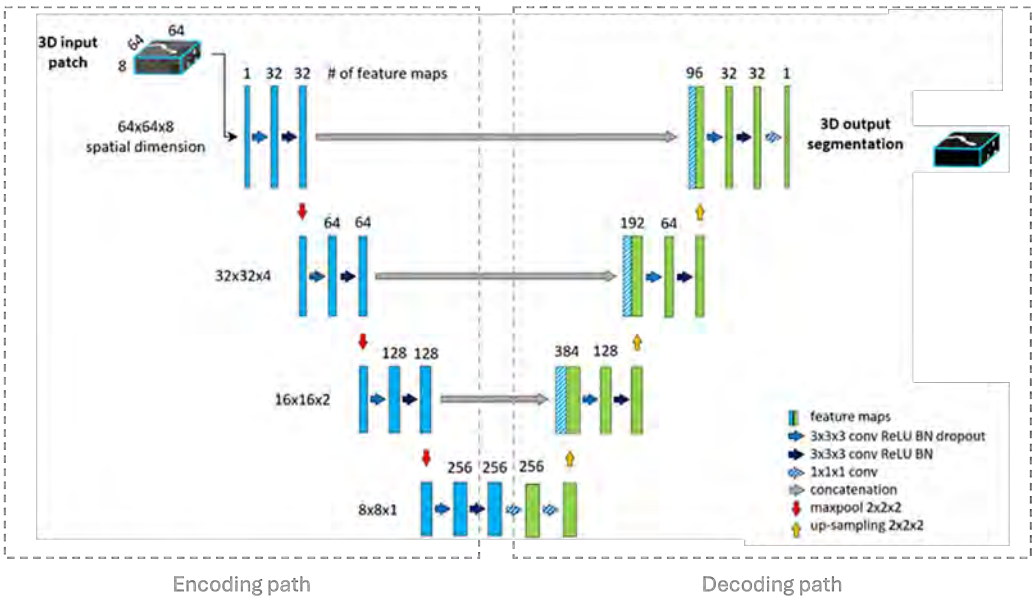


Figure 6.3: Schematic overview of the U-Net architecture. The left leg of the 'U' is the encoding path, responsible for the extraction of meaningful image features. The right leg of the 'U' is the decoding path, which combines all image features maps at different stages of the encoding path into segmentation mask at the same resolution of the original image. Image adapted from [6]

As the spatial dimensions decrease, the information becomes increasingly compressed, which is advantageous for capturing context. However, this could lead to a loss of spatial resolution in the final network output. To counteract this, the U-Net introduces skip connections that directly transfer feature maps at different levels from the encoder path to the corresponding layers in the decoder. These connections allow the decoder to leverage the intermediate high-resolution features learned during the encoding process, while preserving spatial information that would otherwise be lost.

The expansive path, or decoder, then aims to recover the spatial resolution and construct a detailed segmentation map. This path consists of upsampling layers, which are implemented as transposed convolutions. Each upsampling operation doubles the spatial dimensions of the feature maps (as per default, arbitrarily chosen). These are followed by concatenation with the corresponding feature maps from the encoding

path, provided through the skip connections. After concatenation, a series of 3x3 convolutions and an activation function (typically ReLu) refine the upsampled features. The final layer is a 1x1 convolution that reduces the number of feature channels to the desired number of classes for segmentation.

One of the key innovations of U-Net was the use of skip connections, which address the challenge of recovering spatial information lost during the downsampling process. By combining the coarse, abstract features from the deeper layers with the fine, detailed features from the shallower or upper layers, the network can produce segmentation maps with both high accuracy and spatial resolution.

The base U-Net architecture has been adapted and extended in numerous ways to tackle a wide range of image segmentation problems also beyond biomedical applications. Variants such as 3D U-Net have been developed for volumetric data, making it suitable for tasks involving 3D medical imaging modalities like CT and MRI scans. Additionally, various modifications have been introduced to enhance the performance and efficiency of U-Net, including attention mechanisms, deeper and wider networks, and the incorporation of residual connections.

The impact of U-Net on the field of image segmentation and computer vision cannot be overstated. Its simple yet effective design has made it the standard architecture for many segmentation tasks. The principles underlying U-Net, such as the encoder-decoder structure and the use of skip connections, have influenced the design of numerous subsequent architectures in computer vision.

6.4 nnU-Net

While the original U-Net already dates from 2015, its successors and derivates are still state-of-the-art today. The currently most famous and frequently used derivate of the original U-Net is the nnU-Net (also known as 'no new net') [7]. The model name is already indicative for its architecture: it is also just a U-Net, but the innovation lies in its self-adapting nature of the model (hyper)parameters. In a traditional U-Net, all (hyper)parameters and model architecture parameters (number of layers, kernel size of convolutions,...) should be finetuned manually (which is huge time-consuming task). On the other hand, nnU-Net automates these processes, thereby eliminating the need for this massively time-consuming step. Futhermore, the open-source framework is characterized by its ease-of-use and the implementation of multiple useful tools for performing automated segmentation experiments (checkpointing, logging of model parameters, logging of debug information, etc.).

6.4.1 Automated configuration

nnU-Net's automation begins already during the first dataset loading. Upon receiving a new dataset, nnU-Net evaluates its characteristics, such as image size, spacing, and intensity distribution. Based on this analysis, it selects the optimal preprocessing steps, including resampling, normalization, and useful data augmentation techniques. The framework then determines the most suitable network configuration from a predefined

set of U-Net variants, tailored to the dataset's specific needs (Figure 6.4). Notably, the framework not only considers the dataset's specific needs but also takes into account the constraints imposed by available hardware resources. In particular, the size of the GPU memory significantly influences the determination of the image patch size. Consequently, nnU-Net offers a valuable tool for efficiently maximizing the use of all available computational power, eliminating the need for time-consuming trial-and-error experimentation.

6.4.2 Predefined network architectures: 2D and 3D approaches

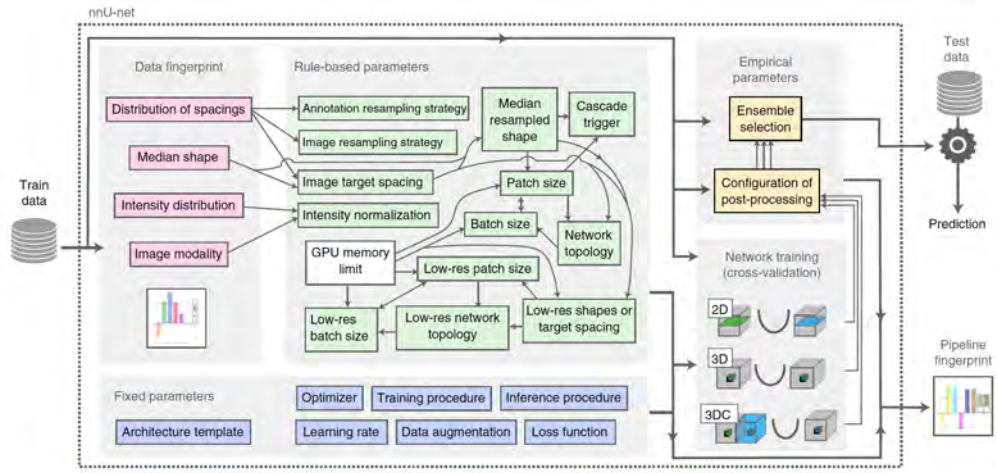
The nnU-Net framework employs three distinct U-Net configurations: a 2D U-Net, a 3D U-Net, and a cascaded 3D U-Net. The 2D U-Net processes slices of 3D scan volumes independently one by one, while the 3D U-Net handles entire 3D volumes at once, providing better spatial context at the cost of increased computational demand (GPU memory and operations). The cascaded 3D U-Net first applies a coarse 3D U-Net to the downsampled image (low-resolution) and then performs a refinement step using a second 3D U-Net with the full-resolution image and upsampled coarse 3D U-Net segmentation results stacked in the input channel dimension.

6.4.3 Training and optimization

nnU-Net automatically employs the best-suited set of hyperparameters for training, such as learning rate, batch size, and loss function, through a series of experiments with cross-validation. It uses advanced optimization techniques, including stochastic gradient descent (SGD) with momentum and adaptive learning rate schedules, to ensure robust training. nnU-Net also incorporates strategies to address common challenges in biomedical image segmentation, such as class imbalance and small object segmentation, by employing techniques like patch sampling and class-specific loss functions. Finally, some post-processing techniques such as small particle removal are also readily available.

6.4.4 Conclusion

The nnUNet represents a significant advancement in the field of biomedical image segmentation. Its automated, self-adapting framework simplifies the development of high-performing segmentation models, democratizing access to advanced deep learning techniques. By consistently achieving state-of-the-art results (or at least very close to state-of-the-art) across diverse tasks and datasets, nnUNet has established itself as a benchmark in the field.



Design choice	Required input	Automated (fixed, rule-based or empirical) configuration derived by distilling expert knowledge (more details in online methods)
Learning rate	–	Poly learning rate schedule (initial, 0.01)
Loss function	–	Dice and cross-entropy
Architecture template	–	Encoder–decoder with skip-connection ('U-Net-like') and instance normalization, leaky ReLU, deep supervision (topology-adapted in inferred parameters)
Optimizer	–	SGD with Nesterov momentum ($\mu = 0.99$)
Data augmentation	–	Rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring
Training procedure	–	1,000 epochs × 250 minibatches, foreground oversampling
Inference procedure	–	Sliding window with half-patch size overlap, Gaussian patch center weighting
Intensity normalization	Modality, intensity distribution	If CT, global dataset percentile clipping & z score with global foreground mean and s.d. Otherwise, z score with per image mean and s.d.
Image resampling strategy	Distribution of spacings	If anisotropic, in-plane with third-order spline, out-of-plane with nearest neighbor Otherwise, third-order spline
Annotation resampling strategy	Distribution of spacings	Convert to one-hot encoding → If anisotropic, in-plane with linear interpolation, out-of-plane with nearest neighbor Otherwise, linear interpolation

Image target spacing	Distribution of spacings	If anisotropic, lowest resolution axis tenth percentile, other axes median. Otherwise, median spacing for each axis. (computed based on spacings found in training cases)
Network topology, patch size, batch size	Median resampled shape, target spacing, GPU memory limit	Initialize the patch size to median image shape and iteratively reduce it while adapting the network topology accordingly until the network can be trained with a batch size of at least 2 given GPU memory constraints. for details see online methods.
Trigger of 3D U-Net cascade	Median resampled image size, patch size	Yes, if patch size of the 3D full resolution U-Net covers less than 12.5% of the median resampled image shape
Configuration of low-resolution 3D U-Net	Low-res target spacing or image shapes, GPU memory limit	Iteratively increase target spacing while reconfiguring patch size, network topology and batch size (as described above) until the configured patch size covers 25% of the median image shape. For details, see online methods.
Configuration of post-processing	Full set of training data and annotations	Treating all foreground classes as one; does all-but-largest-component-suppression increase cross-validation performance? Yes, apply; reiterate for individual classes No, do not apply; reiterate for individual foreground classes
Ensemble selection	Full set of training data and annotations	From 2D U-Net, 3D U-Net or 3D cascade, choose the best model (or combination of two) according to cross-validation performance

Figure 6.4: Schematic overview of the automated configuration of the nnU-Net. Parameter interdependencies are represented by thin arrows and are modelled by a set of heuristic rules. Finally, the data-dependent 'rule-based parameters' are inferred (green). The blue boxes represent 'fixed parameters'. Cross-validation (5-fold) is implemented in the default training configuration. Depending on the dataset, up to three configurations are trained (2D, low-resolution 3D, cascade or high-resolution 3D). Finally, an empirical selection of the best performing ensemble on the available dataset is done, including the decision to perform post-processing (yellow boxes represent the 'empirical parameters'. The table below summarizes how all parameters are configured by the nnU-Net framework. [7]

6.5 Swin-UNetR

Before delving into the Swin-UnetR model architecture, it's essential to learn about some essential characteristics of Transformer models. Initially designed for natural language processing (like GPT to mention the most famous one), these building blocks for neural networks have revolutionized various fields, including computer vision.

6.5.1 Transformers

At the core of a transformer is the self-attention mechanism. Unlike recurrent neural networks (RNNs), that process sequences sequentially (one by one, in a fixed order), transformers can process information from different parts of the input simultaneously. This allows for a more efficient modeling of long-range dependencies. Non-overlapping image patches are first normalized (layer norm), before fed into the multi-head attention network. Based on the similarity between elements, the self-attention mechanism calculates a weighted sum of all input elements for each element. This allows the model to focus on the most relevant parts of the input and capture also global dependencies and patterns.

6.5.2 Swin Transformer

The Swin Transformer [8], a variant of the transformer architecture models, was specifically designed for computer vision tasks. It introduces a hierarchical structure, whose representation is computed with shifted windows (Swin). By alternating between these shifted patch windows in successive layers, cross-window connections are introduced. Similar to other transformer architectures, it divides the input image into non-overlapping patches and progressively reduces the spatial resolution while increasing the channel dimension. This hierarchical structure allows the Swin Transformer to capture both local and global information efficiently.

6.5.3 Swin-UnetR: A Transformer-Based Medical Image Segmentation Model

Swin-UnetR [9] builds further upon the Swin Transformer and U-Net architectures to create a powerful model for medical image segmentation. From U-Net, it employs the encoder-decoder structure to capture both context and fine-grained details (Figure 6.5). However, instead of using convolutional layers, it uses the Swin Transformer blocks to extract the intermediate image feature maps. This allows the model to benefit from the long-range dependencies (shallow layers) and hierarchical structure of the Swin Transformer, while still capturing spatial information effectively.

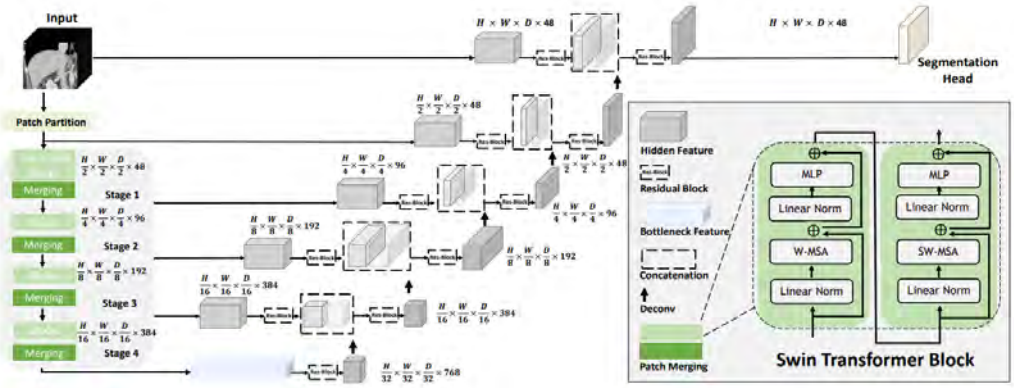


Figure 6.5: Swin-UnetR model architecture [9].

Architectural characteristics of the Swin-UnetR include:

- **Encoder:** The encoder part of Swin-UnetR consists of multiple Swin Transformer blocks. Each block applies self-attention to the input features, followed by a feed-forward network. As we move deeper into the encoder, the spatial resolution is reduced while the channel dimension is increased, allowing the model to capture higher-level features.
- **Decoder:** The decoder part of Swin-UnetR uses a similar architecture to the encoder but with upsampling operations to increase the spatial resolution. The decoder also incorporates skip connections, which combine features from the encoder with those from the decoder, helping to preserve fine-grained details.
- **Output Layer:** The final layer of Swin-UnetR is a convolutional layer that produces a segmentation map, where each pixel is assigned a class label.

The key advantages of Swin-UnetR are the following:

- **Long-range dependencies:** The Swin Transformer blocks enable the model to capture long-range dependencies within the image, which is crucial for medical image segmentation tasks.
- **Hierarchical structure:** The hierarchical structure of the Swin Transformer allows the model to capture both local and global information, improving segmentation accuracy.
- **State-of-the-art performance:** Swin-UnetR has achieved state-of-the-art performance on various medical image segmentation benchmarks, demonstrating its effectiveness.

6.5.4 Conclusion

In conclusion, Swin-UnetR represents again a step forward in medical image segmentation. By combining the strengths of transformers and U-Net, it offers a powerful and efficient approach for tasks such as organ segmentation, lesion detection, and more.

6.6 MedNeXt

While transformers indeed deserve their place amongst the state-of-the-art computer vision models, they still face some difficulties when applied on smaller datasets. Inspired by the hierarchical transformer architectures (such as Swin Transformers, described before), the ConvNeXt network family was developped: a pure convolutional network without actual transformer blocks. Their performance is on par with Transformer networks in terms of accuracy and scalability, while its simplicity and efficiency is inherited from standard convolutional networks. The MedNeXt architecture aims to extend the benefits of the ConvNeXt networks to achieve state-of-the-art performance also for segmentation of biomedical images.

6.6.1 From ConvNeXt...

The following model architecture design improvements were explored in [10]:

- Transformers divide the input image typically in **non-overlapping patches**, before extracting the image features for each patch. The implementation of a 4x4 convolutional layer with a stride of 4 has the same effect.
- By implementing grouped convolutions (number of groups = number of channels), combined with 1x1 convs, **spatial and channel mixing is separated**, similar to vision Transformers.
- Similar to Transformer blocks, the **inverted bottleneck** (hidden layer channel dimension of the MLP block being 4 times its input channel dimension) is implemented in the ConvNeXt blocks.
- By moving up the position of the depthwise convolutional layer before the inverted bottleneck (similar to Transformers where the multi-head attention block is placed before the MLP layers), the more complex large kernel convolutional layers only have to take into account a smaller account of channels (or feature maps). The increase and decrease of the channel dimension in the inverted bottleneck can then be performed by the more efficient 1x1 layers. **The large kernel sizes** allow for a significant increase in receptive field of the convolutional layers, and thus increases the ability to capture longer-distance dependencies. The accuracy improvement however reach a plateau at certain kernel size.

6.6.2 ...to MedNeXt

Originating from the ConvNeXt, the following changes were made to adapt MedNeXt to the more complex taks of biomedical image (volume) analysis [11]:

- **3D imaging:** While ConvNeXt was mainly designed for 2D image tasks, MedNeXt allows for 3D medical images to be processed. All intermediate blocks where the image data flows through are adapted to accept 3D imaging/volume data as input.
- **ConvNeXt blocks:** The further introduction of the inverted bottleneck also in the '2x down' (encoder, blue in Figure 6.6) and '2x up blocks' (decoder, green in Figure 6.6 is one of the changes of MedNeXt in comparison to ConvNeXt.
- **Resampling:** is done by the insertion of a strided (stride=2) convolution or transposed convolution in the first depthwise convolutional layer of the block. The corresponding channel increase or reduction is employed by the last compressive convolutional layer of the block. Finally, a residual connection with 1x1x1 (transposed) convolution with a stride of 2, enables a better gradient flow during backpropagation.

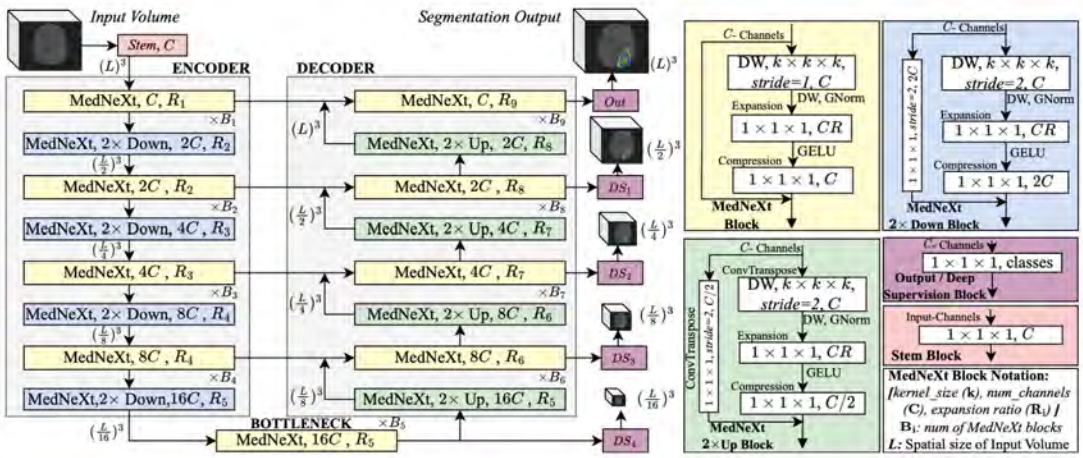


Figure 6.6: Schematic overview of the MedNeXt architecture. [11]

6.7 SegMamba

6.7.1 Introduction

The previously described neural networks were (more or less) ordered following the model's receptive field, in increasing order. The idea is that, similarly to manual segmentation of full scans versus small patches, the model can learn some spatial awareness related to the anatomy of interest, thereby acting as some sort of regularisation. CNN-based methods have difficulties in recognizing global image

relationships, as the convolutional layer acts by definition more on localized features. The next upcoming family of segmentation algorithms, based on the Transformer architecture, are employing a self-attention module to extract global information. Hybrid approaches, combining the benefits of both segmentation algorithm families have become state-of-the-art in a number of segmentation tasks and challenges. The main drawback of implementing transformer architectures into computer vision models, is the increased computation cost, related to the quadratic complexity (O^2) in the self-attention modules. Following the success of the transformers for computer vision tasks, a new family of neural networks have emerged: Mamba-based models originate from state space models and are developed to capture long-range dependencies optimally while an efficiency improvement results in still feasible computational requirements. Multiple derivations from this work in the natural language processing domain have emerged. However, to date Mamba-based models were not fully explored in the domain of MRI knee segmentation. SegMamba [12] is one of the first segmentation algorithms that combines elements from the U-Net architecture (U-shape structure to capture image volume features at different scales) with Mamba blocks, allowing for the segmentation of 3D image volumes.

6.7.2 State space models

For clarity, the term "state space models" will not be abbreviated in this work to prevent confusion with the more commonly used concept of statistical shape model (SSM) troughout this work. State space models were a concept originally developed in the field of control engineering and signal processing [13]. They offer a way to model processes where the driving forces (*state*) behind the variable of interest (*space*) are not measurable. Where recurrent neural networks and convolutional neural networks lack the ability to efficiently select the most informative data from the input, selective state space models such as Mamba-based models have implemented a selection mechanism by parametrizing the state space model parameters based on the input. Therefore, irrelevant information is filtered out and important features are more efficiently captured by the model's state. As neural network model architectures and implementation details are not the scope of this work, the interested reader is referred to the base paper on Mamba [14] and this extensive survey paper on Mamba-based neural networks [15].

6.7.3 Key architecture features of SegMamba

As the basic Mamba-blocks are designed to process one-dimensional input data, SegMamba extends this to the whole-volume sequential modeling of three-dimensional data, by implementing a tri-orientated spatial Mamba (TSMamba) block. It flattens the three-dimensional input in three directions: forward, reverse and inter-slice. Additionally, gated spatial convolution (GSC) modules aims to enhance the spatial feature representation and feature-level uncertainty estimation (FUE) filter the multi-scale features from the encoder, enhancing feature reuse.

Tri-orientated spatial mamba (TSMamba) block

These TSMamba blocks are composed out of a gated spatial convolution, a normalization layer, a tri-orientated Mamba module (ToM), another normalization layer and a MLP layer, with residual connections as illustrated on the left in Figure 6.7.

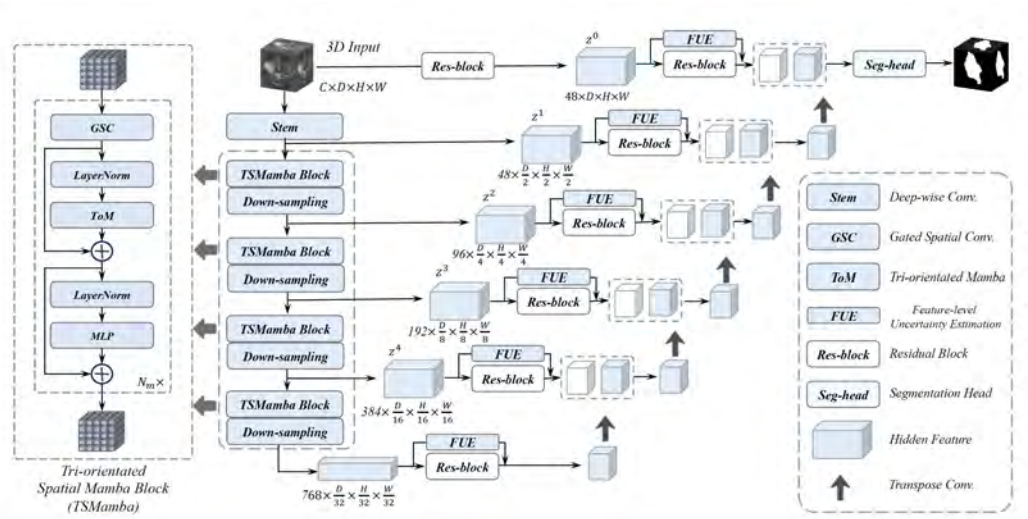


Figure 6.7: An overview of the SegMamba model architecture. The encoder includes a stem layer and multiple TSMamba blocks designed to extract features at multiple scales. Within each TSMamba block, a gated spatial convolution (GSC) module captures the spatial features, and a tri-orientated Mamba (ToM) module extracts global information from forward, backward and inter-slice directions. [12]

Gated spatial convolution (GSC) Where the original Mamba layer omits spatial information during flattening of the 3D features into a 1D sequence, the gated spatial convolution (GSC) module captures the spatial relationships. Two convolutional blocks (containing a norm, convolution, and nonlinear layer) with kernel sizes $3 \times 3 \times 3$ and $1 \times 1 \times 1$ are employed, followed by a pixel-by-pixel multiplication acting like a gate mechanism. The features are finally fused by another convolutional block, as illustrated in Figure 6.8.

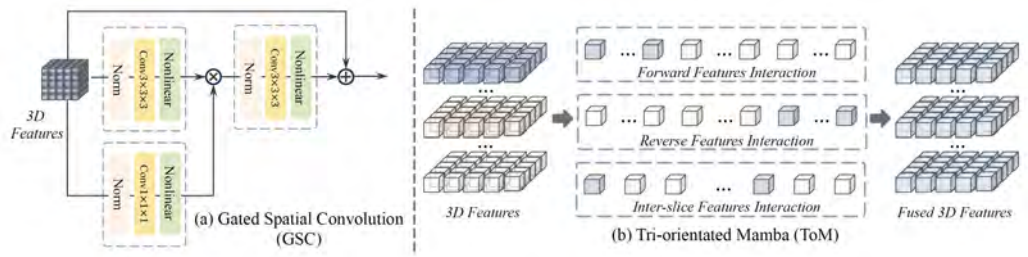


Figure 6.8: (a) The gated spatial convolution. **(b)** The tri-orientated Mamba.[12]

Tri-orientated Mamba module (ToM) The original Mamba block only captures global dependencies in one direction (e.g. forward), thereby neglecting potential dependencies in other directions (e.g. backward, out-of-plane or over multiple slices). To accomodate for the high-dimensional nature of medical scans, the tri-orientated Mamba module computes feature dependencies from three directions: forward, backward and inter-slice. These three types of features are finally fused by summing them up.

Feature-level uncertainty estimation (FUE)

To estimate the feature-level uncertainty, the mean value across the channel dimension is calculated and fed into a sigmoid function σ for normalization purposes. The uncertainty u^i is then calculated as in Eq. 6.1.

$$u^i = -\bar{z}^i \log(\bar{z}^i), \text{ and } \bar{z}^i = \sigma(\frac{1}{C^i} \sum_{c=1}^{C^i} {z^i}_c) \tag{6.1}$$

The final i^{th} scale feature is then computed as follows (Eq. 6.2):

$$\tilde{z}^i = z^i + z^i(1 - u^1) \tag{6.2}$$

6.7.4 Conclusion

The described additions to the Mamba base network enable SegMamba to model long-range dependencies within volumetric data, while still operating at a high level of efficiency during inference.

Bibliography

[1] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al.. PyTorch: An Imperative Style, High-Performance Deep Learning Library; 2019. Available from: https://arxiv.org/abs/1912.01703.

[2] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al.. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. Software available from tensorflow.org. Available from: https://www.tensorflow.org/.

[3] Pramoditha R. The Concept of Artificial Neurons (Perceptrons) in Neural Networks: Neural Networks and Deep Learning Course (Part 1). Medium.com; 2021. Available from: https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc.

[4] Meyer-Baese A, Schmid V. Chapter 7 - Foundations of Neural Networks. In: Meyer-Baese A, Schmid V, editors. Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition). second edition ed. Oxford: Academic Press; 2014. p. 197-243.

[5] Ronneberger O, Fischer P, Brox T. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. U-Net: Convolutional Networks for Biomedical Image Segmentation; 2015. p. 234-41.

[6] Hilbert A, Madai VI, Akay EM, Aydin OU, Behland J, Sobesky J, et al. BRAVE-NET: Fully Automated Arterial Brain Vessel Segmentation in Patients With Cerebrovascular Disease. Frontiers in Artificial Intelligence. 2020 9;3.

[7] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2021 2;18:203-11.

[8] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al.. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows; 2021. Available from: `https://arxiv.org/abs/2103.14030`.

[9] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images; 2022. Available from: `https://arxiv.org/abs/2201.01266`.

[10] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022. p. 11966-76.

[11] Roy S, Koehler G, Ulrich C, Baumgartner M, Petersen J, Isensee F, et al.. MedNeXt: Transformer-driven Scaling of ConvNets for Medical Image Segmentation; 2023. Available from: `https://arxiv.org/abs/2303.09975`.

[12] Xing Z, Ye T, Yang Y, Liu G, Zhu L. SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation; 2024. Available from: `https://arxiv.org/abs/2401.13560`.

[13] Kalman RE. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering. 1960 3;82:35-45.

[14] Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces; 2024. Available from: `https://arxiv.org/abs/2312.00752`.

[15] Zhang H, Zhu Y, Wang D, Zhang L, Chen T, Wang Z, et al. A Survey on Visual Mamba. Applied Sciences. 2024 6;14:5683.

Chapter 7

Automated Landmark Annotation for Morphometric Analysis of Distal Femur and Proximal Tibia

This chapter describes our proposed algorithm for automated landmark annotation in the context of bone morphometric analysis. The manual landmark-based method was validated based on intra- and interobserver variability, and this was in turn compared to the intermethod variability. The achieved high level of automation will lead to a faster, scalable and human operator-independent morphometric analysis of the knee bones.

This work was previously published as: *Grammens, J., Van Haver, A., Lumban-Gaol, I., Danckaers, F., Verdonk, P., Sijbers, J. (2024). Automated Landmark Annotation for Morphometric Analysis of Distal Femur and Proximal Tibia. Journal of imaging, 10(4), 90. https://doi.org/10.3390/jimaging10040090*

7.1 Introduction

In orthopedic clinical practice, patient-specific instruments and implants are highly dependent on accurate anatomical characterization for the region of interest, for which landmark analysis is a commonly used technique. It involves the use of 3D coordinates of anatomically meaningful points to calculate relative distances and angles. Several knee-specific applications exist, including a sizing and positioning tool for joint replacement surgery [1], a matching tool for donors and acceptors in meniscal allograft transplantation [2], or large-scale morphometric risk factor analysis [3, 4, 5]. All these applications are based on a set of anatomical landmarks at the level of the distal femur and proximal tibia. Traditional methods to identify anatomical landmarks involve manual annotation, which is highly time-consuming and prone to substantial intra- and inter-observer variability [6, 7]. Time consumption depends mainly on image quality and experience level, plateauing towards the end of the learning curve. Clear landmark definitions and well-described protocols are essential to limit this human operator-induced variability [8]. Various automated approaches have been developed to minimize subjectivity and processing time in annotation tasks. They can be divided

into three large groups: knowledge-based, template-based and learning-based [9]. The first category exploits knowledge of the human anatomy to characterize landmarks by a set of unique geometric or anatomical characteristics such as surface normal direction, local curvature, or spatial relation to other landmarks [10]. Template-based approaches like statistical shape models start with an annotated template image or template mesh, which is elastically deformed to match the image or mesh of the subject's anatomy [11]. Lastly, several deep learning-based approaches have been proposed [12, 13], mainly consisting of combinations or modifications to convolutional neural networks. We propose a hybrid methodology combining a template-based algorithm with knowledge-based optimization of landmark positions. More specifically, a fully automated landmark annotation method is proposed for measurement of the medial and lateral tibiofemoral joints in all three directions. Anatomical correspondences are established by 3D surface registration [11], after which the landmark seeds are propagated from the template mesh to the subject's 3D bone and cartilage surface meshes. Finally, landmarks defined at local extreme positions were optimized to ensure their position at that outermost point. Our proposed automated landmarking tool is compared with the average landmark coordinates and measurements from three independent expert observers. Furthermore, the landmarking tool is tested for the 3D morphometric analysis of the tibiofemoral joint, based on a retrospectively collected dataset of 3D models of 20 knees. It does so by comparing the intra- and inter-observer variability (manual, within, and between observers) with the inter-method (manual versus automated method) variability and reliability. We hypothesize the inter-method landmark variability to be non-inferior to the inter-observer landmark positioning variability. Furthermore, we hypothesize that our automated morphometric measurement variability is in line with the inter-observer morphometric measurement variability.

7.2 Methods

7.2.1 Data and workflow

From a large multicenter database of arthroscopic partial meniscectomy patients, 20 randomly selected anonymized pre-operative knee MRI scans were randomly selected: 15 male (mean age ± SD: 49 ± 13 years) and 5 female (mean age ± SD: 58 ± 13 years) subjects. Subjects with osteophytes (OARSI grade I–III) or major knee deformities (>5° varus or valgus by clinical judgement) were excluded from the study. Informed consent was obtained from all participants prior to their inclusion and the study protocol was approved by the ethical committees according to the 1964 Declaration of Helsinki and its later amendments.

Manual image segmentation The MRI scan data were loaded into Mimics 22.0 (Materialise, Leuven, Belgium) to create 3D models of the distal femur and proximal tibia. For all subjects, MRI image sets in the three perpendicular anatomical planes were available, following the standard clinical MRI protocols. At least two MRI image sets with a perpendicular acquisition plane were used to segment the distal femur and proximal tibia into two separate 3D models consisting of bone and cartilage. The

resulting projected contours of the 3D models were then visually double-checked on the remaining available image sets and manually finetuned in an iterative workflow using the "Contour edit" tool of the software package. Finally, the 3D models of the distal femur and proximal tibia were saved as triangular 3D surface meshes.

7.2.2 Manual landmark annotation

All included subjects (n = 20) were manually landmark-annotated by three trained observers: one junior researcher (JG), one senior researcher (AVH), and one orthopedic surgeon (ILG). All observers had at least 2 years of experience. One observer (JG) performed all landmark annotations three times in a random case order with a minimum interval of 1 month to avoid recall bias. The landmarks and morphometric measurements of interest are defined and further described in '3.Observations'.

7.2.3 Automated landmark annotation

Registration of 3D Bone and Cartilage Models

Three-dimensional surface registration was performed to obtain a dense set of anatomically corresponding pseudo-landmarks. The first step consisted of isotropic remeshing of an arbitrary chosen bone and cartilage shape of interest [14], which served as a template. Next, an iterative algorithm of rigid and constrained elastic deformations [11] was used to register all 3D bone and cartilage meshes with the template (see also Section 11.3. Surface registration). Next, the mean distal femur and proximal tibia bone and cartilage shape were calculated by averaging the corresponding point coordinates of the deformed meshes. Using the same iterative registration process as described before, all 20 bone and cartilage shapes were again registered with the mean bone and cartilage shape as a template. The mean distal femur and proximal tibia meshes consisted of 47,622 and 46,721 vertices, respectively. All resulting deformed meshes were visually checked for resulting mesh quality (triangle distortion).

Landmark propagation

From the obtained anatomical correspondences, all manually annotated landmarks (cf. 9.2.2 Manual landmark annotation) of all but one subject (the subject of interest for automated landmark annotation) were propagated to the mean bone and cartilage shape in a leave-one-out experiment. Their coordinates were averaged to define the mean landmark position, projected on the mean bone and cartilage shape. To ensure equal weighting of the three observers in the mean of the landmark projections, the three landmark projections of observer 1 were averaged beforehand. Next, automated landmark annotation of all subjects was initialized by propagating the mean landmark positions to the corresponding vertices of the subjects. In a final step, the landmark positions at extreme locations were optimized according to their definition (9.2.4 Landmark definitions) via a custom Python script (and VTK library [15]) to ensure their

position was the most anterior/posterior, medial/lateral or proximal/distal point. The complete workflow is summarized in Figure 9.1.

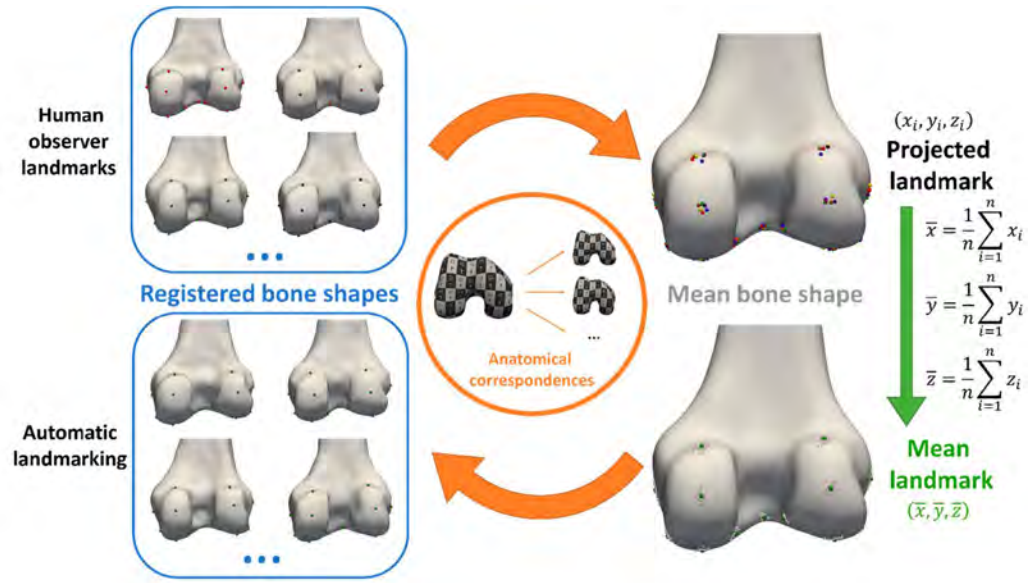


Figure 7.1: Workflow overview. Landmarks were annotated manually by three experts **(top left)** and propagated to the mean bone and cartilage shape **(top right)**, leaving out the observations of the subject for automated landmarking. Next, mean landmark coordinates were calculated from all propagated observations on the mean bone and cartilage shape **(bottom right)**. Finally, automated landmarking initialization was performed by propagating the landmarks back to the subject's bone and cartilage shapes, using the previously established anatomical correspondences again **(bottom left)**

7.2.4 Observations

Landmark definitions

The landmark definitions were adopted from earlier reported landmark studies [3, 6, 5, 7]. The landmarks were manually annotated in 3D Slicer, first directly on the 3D model and eventually finetuned on the relevant MRI view with projected 3D model contours. A complete overview of the evaluated landmarks with their acronyms and definitions is given in Table 9.1 and illustrated in Figure 9.2.

Table 7.1: Landmark definitions

Acronym	Landmark	Definition
FME	Femoral medial epicondyle	The most anterior and distal osseous prominence over the medial aspect of the 3D distal femur
FLE	Femoral lateral epicondyle	The most anterior and distal osseous prominence over the lateral aspect of the 3D distal femur
FMCP	Femoral medial condyle posterior	The most posterior point of the medial condyle on the 3D femur.
FLCP	Femoral lateral condyle posterior	The most posterior point of the lateral condyle on the 3D femur.
FMCD	Femoral medial condyle distal	The most distal point of the medial condyle on the 3D femur.
FLCD	Femoral lateral condyle distal	The most distal point of the lateral condyle on the 3D femur.
FMTA	Femoral medial trochlea anterior	The most anterior point of the medial trochlea on the 3D femur.
FLTA	Femoral lateral trochlea anterior	The most anterior point of the lateral trochlea on the 3D femur.
FMCPP	Femoral medial condyle posterior proximal	The most proximal point of the cartilage at the posterior medial condyle on the 3D femur. Verified on a sagittal MRI view.
FLCPP	Femoral lateral condyle posterior proximal	The most proximal point of the cartilage at the posterior lateral condyle on the 3D femur. Verified on a sagittal MRI view.
Notch	Femoral notch	The most anterior point in the middle of the femoral notch on a caudal to cranial view of the 3D femur.
FMCIP	Femoral medial condyle internal point	The most lateral point of the cartilage of the medial condyle on a caudal to cranial view of the 3D femur, at the level of one third of the notch depth anteroposteriorly. Verified on a coronal MRI view.
FMCEP	Femoral medial condyle external point	The most medial point of the cartilage of the medial condyle on a caudal to cranial view of the 3D femur, at the level of one third of the notch depth anteroposteriorly. Verified on a coronal MRI view.
FLCIP	Femoral lateral condyle internal point	The most medial point of the cartilage of the lateral condyle on a caudal to cranial view of the 3D femur, at the level of one third of the notch depth anteroposteriorly. Verified on a coronal MRI view.
FLCEP	Femoral lateral condyle external point	The most lateral point of the cartilage of the lateral condyle on a caudal to cranial view of the 3D femur, at the level of one third of the notch depth anteroposteriorly. Verified on a coronal MRI view.
TMIE	Tibial medial intercondylar eminence	The most proximal or highest point of the medial intercondylar eminence.
TLIE	Tibial lateral intercondylar eminence	The most proximal or highest point of the lateral intercondylar eminence.
TMCP	Tibial medial condyle posterior	The most posterior and lateral point of the medial compartment on the 3D tibia. Verified on a sagittal MRI view.
TLCP	Tibial lateral condyle posterior	The most posterior and medial point of the lateral compartment on the 3D tibia. Verified on a sagittal MRI view.
TMCM	Tibial medial condyle medial	The most medial point of the tibial plateau on the 3D tibia, axially aligned following the posterior condylar line of the corresponding femur.
TLCL	Tibial lateral condyle lateral	The most lateral point of the tibial plateau on the 3D tibia, axially aligned following the posterior condylar line of the corresponding femur.
TMCA	Tibial medial condyle anterior	The most anterior point on the cartilage of the medial tibial plateau (on a sagittal MRI view).
TLCA	Tibial lateral condyle anterior	The most anterior point on the cartilage of the lateral tibial plateau (on a sagittal MRI view).

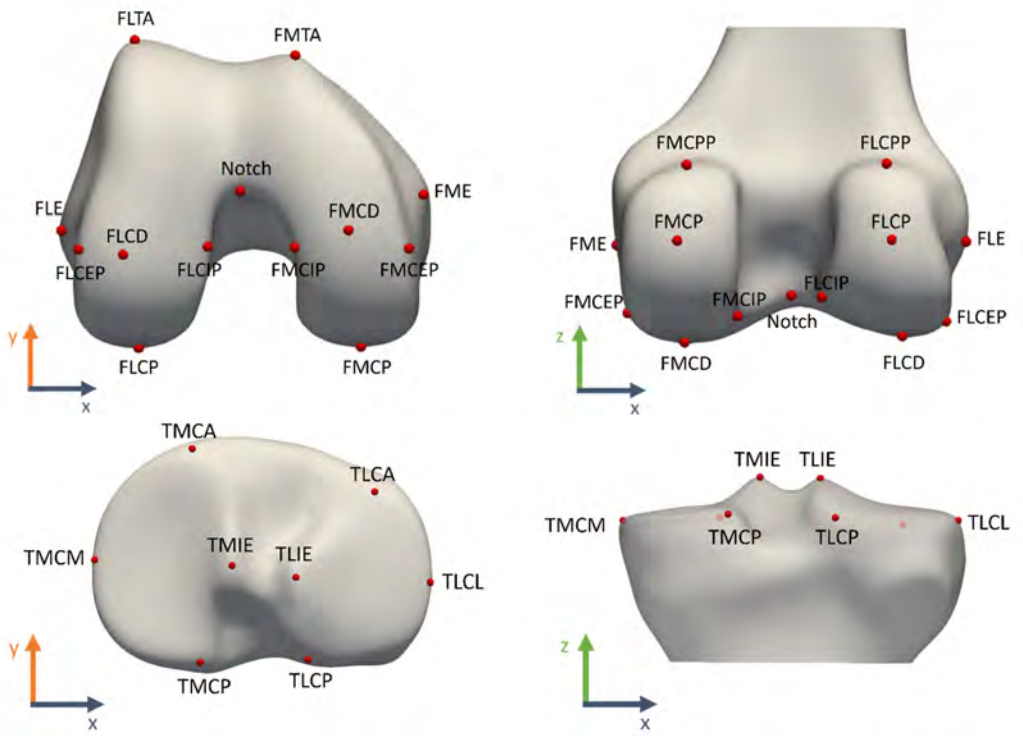


Figure 7.2: Landmark visualization on the template bone and cartilage shapes: **(a)** distal femoral view, **(b)** posterior femoral view, **(c)** proximal tibial view, **(d)** posterior tibial view. A complete overview of landmark definitions can be found in Table 9.1

Reference coordinate system definition

Clinical MRI scans are characterized by a certain field of view, limited to the distal femur and proximal tibia. It is assumed that all knees were correctly positioned by the MRI operators: horizontal and straight on the MRI patient table, leaving only one degree of freedom for minor internal or external rotation of the hip during image acquisition. Therefore, all 3D models of the femur and tibia were rotated in the axial plane to make the femoral posterior condylar line parallel to the mediolateral x-axis. The reference coordinate system is thus defined by the following axes:

- x-axis (mediolateral): parallel to the femoral posterior condylar line, defined by the FMCP and FLCP landmarks (mean of 3 observers as ground truth)
- y-axis (anteroposterior): common perpendicular to the x- and z-axes
- z-axis (proximodistal): MRI patient table movement direction

Morphometric measurement definitions

The morphometric measurements for validation were summarized in Table 9.2 and visualized in Figure 9.3. They characterize the size of the medial and lateral compartments of both the femur and tibia in all clinically relevant directions by projecting the landmarks of interest to the relevant axis.

Table 7.2: Measurement definitions

Measurement Abbreviation	Measurement Definition	Between Landmarks	Measurement Projection Axis
AP MFC	Anteroposterior size of the medial femoral condyle	FMCP, FMTA	y (AP)
AP LFC	Anteroposterior size of the lateral femoral condyle	FLCP, FLTA	y (AP)
AP Notch	Anteroposterior size of the femoral notch	FMCP, Notch (FLCP, Notch)	y (AP)
fML	Mediolateral size of the distal femur	FME, FLE	x (ML)
ML MFC	Mediolateral size of the medial femoral condyle	FMCIP, FMCEP	x (ML)
ML LFC	Mediolateral size of the lateral femoral condyle	FLCIP, FLCEP	x (ML)
ML Notch	Mediolateral size of the femoral notch	FMCIP, FLCIP	x (ML)
PCL	Posterior condylar line	FMCP, FLCP	x (ML)
PD MFC	Proximodistal size of the medial femoral condyle	FMCPP, FMCD	z (PD)
PD LFC	Proximodistal size of the lateral femoral condyle	FLCPP, FLCD	z (PD)
AP MTP	Anteroposterior size of the medial tibial plateau	TMCP, TMCA	y (AP)
AP LTP	Anteroposterior size of the lateral tibial plateau	TLCP, TLCA	y (AP)
tML	Mediolateral size of the tibial plateau	TMCM, TLCL	x (ML)
ML MTP	Mediolateral size of the medial tibial plateau	TMCM, TMIE	x (ML)
ML LTP	Mediolateral size of the lateral tibial plateau	TLCL, TLIE	x (ML)

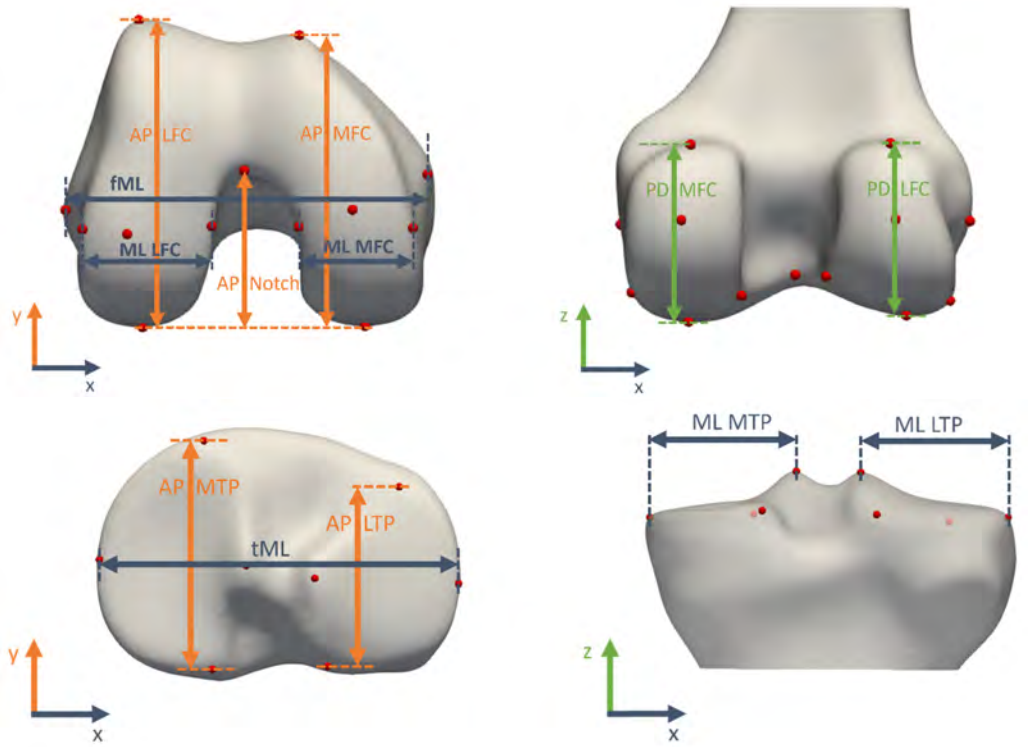


Figure 7.3: Morphometric measurements visualized on the template bone and cartilage shapes: **(a)** distal femoral view, **(b)** posterior femoral view, **(c)** proximal tibial view, **(d)** posterior tibial view. A complete overview of morphometric measurement definitions can be found in Table 9.2

7.2.5 Validation study for manual morphometric analysis

Landmark validation

Manual landmark annotations were considered the gold standard. For the intra-observer error assessment, ground truth landmarks were defined as the average landmark coordinates over three observations of observer 1 and were also used as final landmark annotations for observer 1. For the inter-observer assessment, the ground truth landmarks were defined as the average landmark coordinates over the three observers and further referred to as the expert mean landmarks.

Measurement validation

Alignment to the reference coordinate system was performed using the expert mean landmarks. All measurements were calculated according to their definition in the previous section. For intra-observer variability assessment, the mean of three

measurements by observer 1 served as the ground truth whereas for inter-observer variability evaluation, the mean measurement from three different observers served as the ground truth and is further referred to as the expert mean measurement. Absolute differences with the ground truth were calculated as a measure of variability.

7.2.6 Validation study for automated morphometric analysis

Automated landmark validation

Expert mean landmarks served as ground truth landmark positions. Euclidean distances between the automatically determined landmarks and the ground truth landmarks were calculated as a measure of inter-method landmark variability.

Automated measurement validation

Expert mean measurements served as ground truth morphometric measurements. Automated measurements were calculated from the automated landmark coordinates. Absolute differences were used as a measure of inter-method measurement variability.

7.2.7 Time consumption

The required time to annotate all femoral and tibial landmarks was tracked for one observer and five subject cases. The time needed for the surface registration, extraction of the landmark positions and measurement calculations was derived from the filesystem metadata.

7.2.8 Statistical analysis

All statistical analyses were performed in R 4.2.1, using the 'irr' package for ICC calculations [16, 17].

Landmark postions

The first quartile (Q1), median (Q2), and third quartile (Q3) of the Euclidean distance to the ground truth landmark position were calculated per landmark for intra-observer, inter-observer, and inter-method variability. The interquartile range (IQR = Q3 – Q1) was used to define outliers. Observations below Q1 – 1.5 x IQR or above Q3 + 1.5 x IQR were considered outliers (represented as '+' in the box-and-whisker diagrams of Figure 9.4). The mean and standard deviation of the difference between each observation and the ground truth are calculated per landmark.

Measurements

Similarly, quartiles and medians were calculated per measurement for intra-observer, inter-observer, and inter-method variability. Outliers were again defined as observations that fall below Q1 − 1.5 x IQR or above Q3 + 1.5 x IQR and visualized by '+' in the box-and-whisker diagrams of Figure 9.5. The mean and standard deviation of the absolute differences between each measurement and the mean measurement (intra-observer: three observations of observer 1; inter-observer: three observers) were calculated. Measurement reliability analysis included intra- and inter-observer reliability, reported as intraclass correlation coefficients (ICC) from two-way mixed (intra-observer) or two-way random (inter-observer) effects, absolute agreement, and single-rater models [18]. Values below 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and above 0.9 were, respectively, considered to have poor, moderate, good, and excellent reliability. Inter-method (manual versus automated) agreement was also assessed by the ICC (two-way random effects, absolute agreement, single measurement). Finally, Bland–Altman plots visualized the inter-method (manual versus automated) agreement for all measurements [19].

7.3 Results

7.3.1 Validation study for manual morphometric analysis

Manual Landmark Position Validation

The median landmark position intra-observer difference with the ground truth varied between 0.35 mm (TLIE) and 1.49 mm (TMPA). Maximal intra-observer landmark position differences were between 1.11 mm (TLIE) and 7.66 mm (FME). The median inter-observer difference was between 0.52 mm (TLIE) and 2.28 mm (TLPP). Maximal inter-observer landmark positions ranged between 1.24 mm (TLIE) and 9.7 mm (TLPP). The mean (and standard deviation) of inter-observer landmark differences was 1.53 (±1.22) mm. A complete overview is plotted in a box-and-whiskers diagram in Figure 9.4 (intra-observer: red; inter-observer: green).

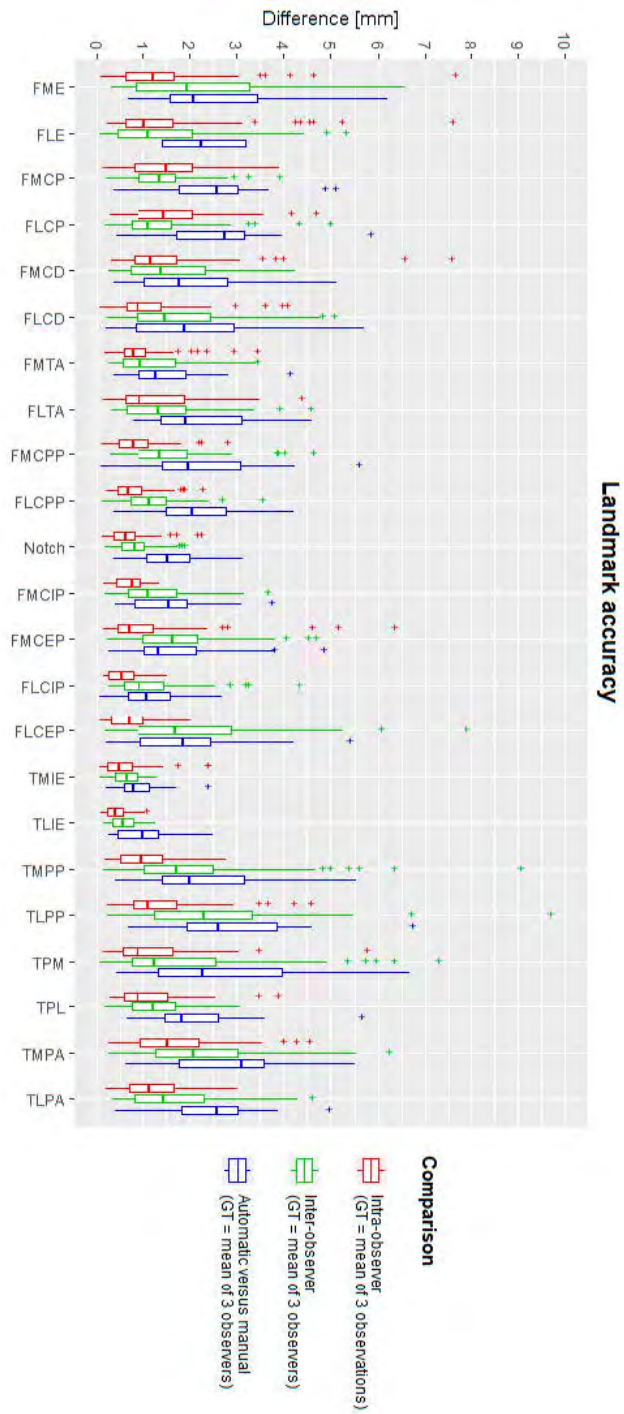


Figure 7.4: Box-and-whisker diagrams for intra-observer, inter-observer and inter-method landmark differences, calculated as Euclidean distances from the ground truth. The boxes indicate the IQR, the line within stands for the median and the whiskers indicate points < 1.5 IQR from the box.'+' represents outliers. IQR: interquartile range, between first and third quartile. GT: ground truth.

Manual Measurement Validation

The median intra-observer measurement differences with the ground truth varied between 0.16 mm (fML) and 0.59 mm (AP LTP). Maximal intra-observer measurement differences ranged between 0.78 mm (AP MFC) and 2.76 mm (AP LTP). The median inter-observer differences ranged between 0.13 mm (AP MFC) and 1.06 mm (AP MTP), whereas maximal differences ranged between 0.83 mm (AP MFC) and 3.02 mm (AP LTP). The mean (and standard deviation) over all inter-observer differences was 0.56 (±0.55) mm. The spread of measurement intra-observer (red) and inter-observer (green) differences is plotted in box-and-whisker plots in Figure 9.5. Intra-class correlation coefficients for intra- and inter-observer errors were between 0.796 and 1. These are summarized in Table 9.3.

Table 7.3: Intraclass correlation coefficients for reliability analysis of the measurements.

Measurement	ICC_{intra}	ICC_{inter}
AP MFC	1	1
AP LFC	1	1
AP Notch	1	1
fML	0.991	0.969
ML MFC	0.937	0.796
ML LFC	0.985	0.922
PD MFC	0.983	0.947
PD LFC	0.976	0.958
AP MTP	1	0.999
AP LTP	0.999	0.998
tML	0.986	0.968
ML MTP	0.953	0.94
ML LTP	0.955	0.935

7.3.2 Validation study for automated morphometric analysis

Automated Landmark Position Validation

The median difference from the ground truth landmark position ranged between 0.77 mm (TMIE) and 3.13 mm (TLPA). The maximal inter-method difference varies between 2.02 mm (FLCIP) and 6.99 mm (TPM). For all landmarks, box-and-whisker diagrams are plotted in Figure 9.4 (blue), and descriptive statistics are in Table 9.4. Success detection rates at different accuracy thresholds are plotted in Figure 9.6. For all landmarks, at least 75% of the automated landmarks were placed within 4 mm of the expert mean landmark. On average, 78% and 92% of the landmarks were placed automatically within 3 mm and 4 mm of the expert mean landmark, respectively.

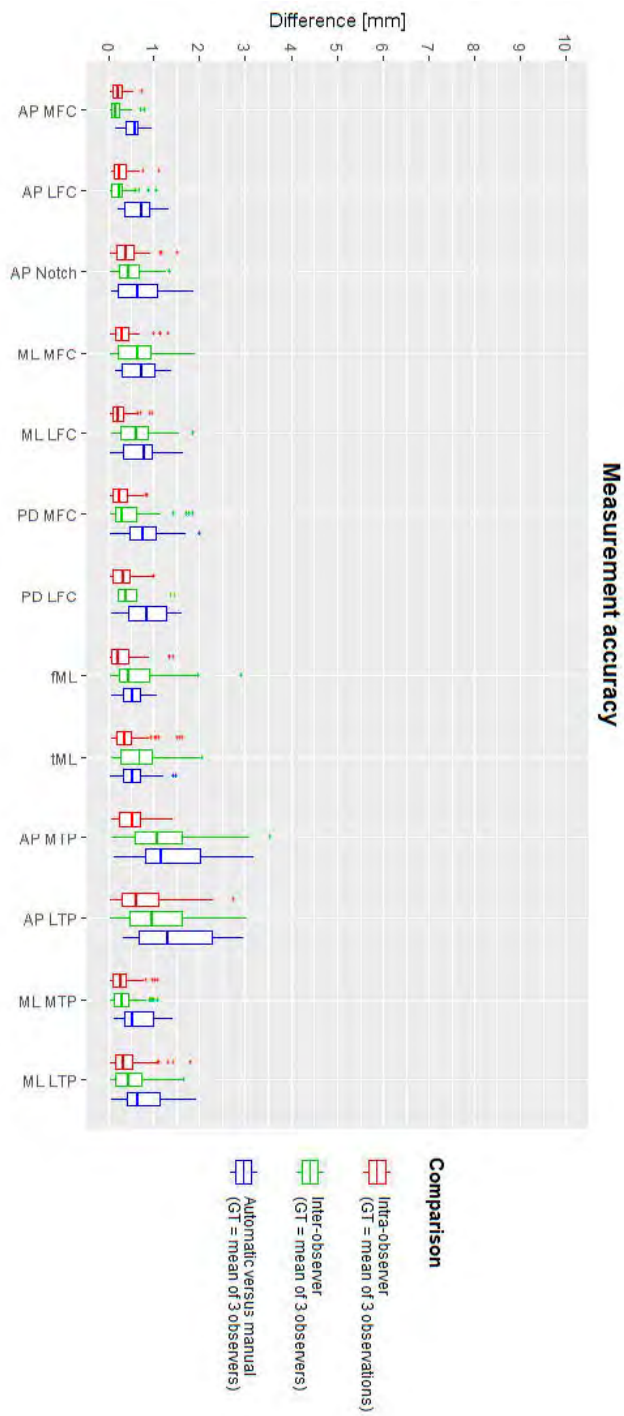


Figure 7.5: Box-and-whisker diagrams for intra-observer, inter-observer, and inter-method (manual versus automated) measurement differences. The mean measurement of three observations (intra-observer) or mean of three observers (inter-observer and inter-method) served as ground truth. The boxes indicate the IQR, the line within stands for the median and the whiskers indicate points < 1.5 IQR from the box. '+' represents outliers. IQR: interquartile range, between first and third quartile. GT: ground truth.

Table 7.4: Descriptive statistics for intra-observer, inter-observer and inter-method (manual versus automated) landmark position differences. Only for the intra-observer differences the mean of three observations was used as reference landmark position. Expert mean landmarks were used as reference landmark positions to assess inter-observer and inter-method landmark differences. SD: standard deviation.

Landmark acronym	Mean (SD) Intra-Observer	Mean (SD) Inter-Observer	Mean (SD) Inter-Method
FME	1.43 (1.26)	2.21 (1.59)	2.59 (1.48)
FLE	1.52 (1.45)	1.47 (1.28)	2.31 (1.07)
FMCP	1.52 (0.91)	1.40 (0.77)	2.51 (1.23)
FLCP	1.59 (1.02)	1.33 (0.95)	2.54 (1.28)
FMCD	1.56 (1.33)	1.56 (1.02)	1.98 (1.26)
FLCD	1.14 (0.87)	1.85 (1.30)	2.06 (1.54)
FMTA	0.91 (0.65)	1.21 (0.83)	1.47 (0.91)
FLTA	1.23 (0.90)	1.44 (0.98)	2.25 (1.20)
FMCPP	0.85 (0.56)	1.55 (0.96)	2.25 (1.34)
FLCPP	0.78 (0.49)	1.16 (0.65)	2.14 (0.99)
Notch	0.69 (0.47)	0.82 (0.46)	1.51 (0.66)
FMCIP	0.69 (0.35)	1.26 (0.78)	1.56 (0.88)
FMCEP	1.05 (1.18)	1.74 (1.03)	1.70 (1.22)
FLCIP	0.54 (0.32)	1.13 (0.84)	1.08 (0.70)
FLCEP	0.70 (0.50)	2.08 (1.59)	1.93 (1.25)
TMIE	0.57 (0.49)	0.62 (0.30)	0.90 (0.53)
TLIE	0.41 (0.25)	0.58 (0.30)	1.05 (0.68)
TMCP	0.99 (0.60)	2.13 (1.67)	2.29 (1.33)
TLCP	1.38 (0.94)	2.58 (1.76)	2.86 (1.45)
TMCM	1.19 (1.01)	1.87 1.72)	2.75 1.88)
TLCL	1.12 (0.76)	1.23 (0.61)	2.13 (1.18)
TMCA	1.64 (0.98)	2.31 (1.34)	2.84 (1.39)
TLCA	1.21 (0.66)	1.63 (1.00)	2.41 (1.15)
All landmarks (average)	**1.07 (0.92)**	**1.53 (1.22)**	**2.05 (1.30)**

Automated Measurement Validation

The median difference between the automated and manual methods was between 0.33 mm (AP MFC) and 1.72 mm (AP MTP). Maximal differences ranged between 0.74 mm (AP MFC) and 2.85 mm (AP MTP). A detailed overview per measurement is visualized as a box-and-whiskers diagram in Figure 9.5, and descriptive statistics are in Table 9.5. Intraclass correlation coefficients varied between 0.938 and 0.999 and are reported in Table 9.6. Success measurement rates are plotted in Figure 9.7 for different accuracy thresholds. For all but two measurements (AP MTP and AP LTP), at least 90% of the

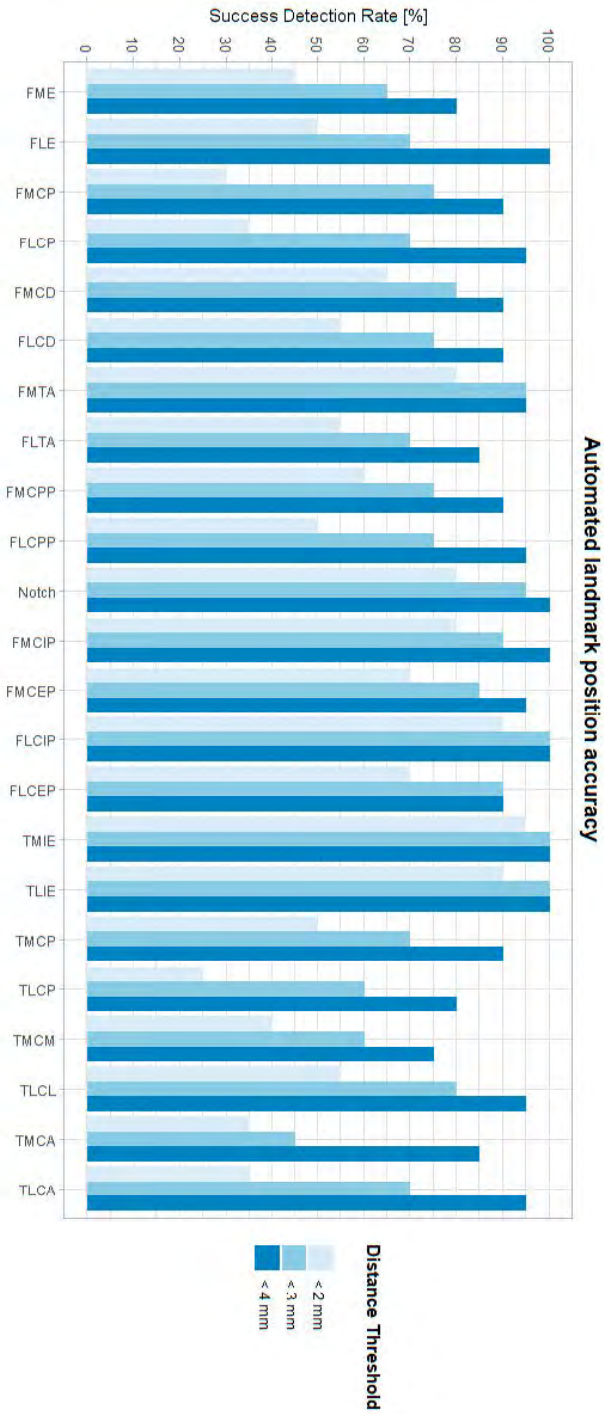


Figure 7.6: Success detection rates in % per landmark within predefined tolerance [mm].

automated measurements were less than 2 mm different from the expert mean measurements. On average, over all measurements, 71% and 95% of all measurements had a difference below 1 mm and 2 mm with the expert mean measurement, respectively. Finally, Bland–Altman diagrams (Figure A.1 in the Appendix chapter) showed similarly distributed errors over the full measurement range.

Table 7.5: Descriptive statistics for measurement differences intra-observer, inter-observer, and intermethod (manual versus automated). Only for the intra-observer differences the mean of three observations was used as reference landmark measurement. Expert mean landmarks were used to calculate the reference measurements to assess inter-observer and inter-method landmark differences. SD: standard deviation.

Measurement	Mean (SD) Intra-Observer	Mean (SD) Inter-Observer	Mean (SD) Inter-Method
AP MFC	0.21 (0.16)	0.18 (0.17)	0.54 (0.21)
AP LFC	0.28 (0.22)	0.23 (0.22)	0.67 (0.36)
AP Notch	0.40 (0.31)	0.46 (0.31)	0.69 (0.53)
fML	0.30 (0.35)	0.63 (0.60)	0.51 (0.28)
ML MFC	0.33 (0.28)	0.61 (0.42)	0.66 (0.41)
ML LFC	0.24 (0.22)	0.62 (0.43)	0.69 (0.45)
PD MFC	0.29 (0.22)	0.45 (0.45)	0.80 (0.51)
PD LFC	0.33 (0.25)	0.43 (0.32)	0.85 (0.51)
AP MTP	0.50 (0.33)	1.19 (0.77)	1.39 (0.91)
AP LTP	0.74 (0.64)	1.11 (0.81)	1.46 (0.92)
tML	0.43 (0.38)	0.70 (0.47)	0.57 (0.45)
ML MTP	0.30 (0.27)	0.35 (0.30)	0.65 (0.41)
ML LTP	0.42 (0.40)	0.50 (0.42)	0.75 (0.57)
All measurements (average)	**0.36 (0.35)**	**0.56 (0.55)**	**0.78 (0.60)**

Table 7.6: Intraclass correlation coefficients (manual versus automated) for reliability analysis of the measurements.

Measurement	ICC	Measurement	ICC
AP MFC	1	AP MTP	0.999
AP LFC	1	AP LTP	0.999
AP Notch	1		
fML	0.995	tML	0.993
ML MFC	0.926	ML MTP	0.944
ML LFC	0.966	ML LTP	0.956
PD MFC	0.961		
PD LFC	0.951		

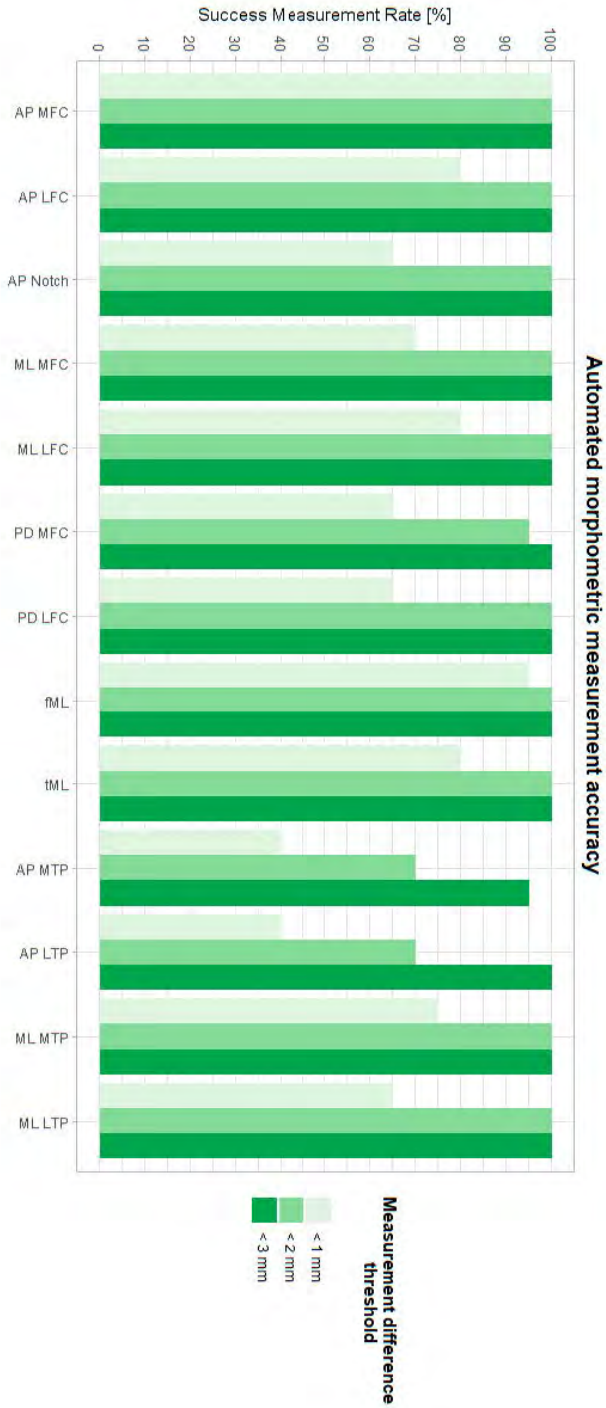


Figure 7.7: Success measurement rates in % per morphometric measurement within predefined tolerance [mm].

7.3.3 Time consumption

The manual method takes 15–30 min per case for extracting all anatomical landmark coordinates from 3D bone and cartilage shapes. The time varied based on observer experience and image resolution. For one observer and five subject cases, the mean (and standard deviation) of the required time for manual landmarking was 18.6 ± 1.8 min per knee. The automated method needed ca. seven minutes of computing time per knee on a desktop workstation (Intel i9-9900K with 32 GB of RAM). This can be split up into the following steps: approx. six minutes for femur and tibia registration, less than one second for initial landmark retrieval and a couple of seconds for landmark optimization.

7.4 Discussion

This study successfully validated our suggested approach for automated anatomical landmarking in 3D morphometric analysis of the distal femur and proximal tibia bones. On average, automated landmarks were placed 2.05 mm from the expert mean landmarks, in comparison to a manual inter-observer variability of 1.53 mm. The derived measurements showed a mean absolute difference of 0.78 mm with the expert mean measurements, whereas the mean inter-observer difference was 0.56 mm. Reliability was proven to be excellent for 3D morphometric measurement of the distal femur and proximal tibia, with an ICC (manual versus automated) ranging between 0.926 and 1. Most importantly, the automated landmark extraction algorithm significantly accelerates the process, decreasing manual labor from approximately half an hour of manual work to a mere seven minutes of operator-independent computing time. Our manual landmark annotations are in line with reported accuracies for the same anatomy. Victor et al. [7] previously reported mean intra-observer differences between 0.41 mm and 1.4 mm and mean inter-observer accuracies between 0.66 mm and 3.5 mm for a highly overlapping set of anatomical landmarks. It should be noted that their 3D bone models were CT-derived (1.25 mm axial slice thickness) and thus potentially had a finer mesh resolution. Van der Merwe et al. [6] obtained mean intra- and inter-observer accuracies in ranges between 0.34 mm and 1.7 mm (intra) and between 0.08 mm and 1.91 mm (inter) on MRI-derived bone models (research scan protocol, 1.5 mm slice thickness, average in-plane resolution of 0.4 mm) of distal femur and proximal tibia. Recently, a method for the assessment of the full lower limb alignment was proposed and validated on CT-derived 3D bone models of the femur and tibia [20]. Based on a similar technology and applied to a largely overlapping set of landmarks, the authors were able to extract all landmarks with a mean absolute difference of 2.17 mm compared to the manual method. Primarily using identical anatomical landmarks, our results are in line with their reported accuracies between repeated manual annotations and inter-method (manual versus automated) accuracy. Several factors may impact the accuracy and success rate of automated landmarking. First and foremost, the quality of the expert landmarks plays a pivotal role in establishing the ground truth. Given the limited number of experts and significant inter-observer variability for certain landmarks, it can be argued if the expert mean landmarks truly serve as a robust ground truth [10]. Interobserver dispersion of the landmarks is likely influenced by the difficulty of manually locating the landmark [21] and variations in training or background among observers [22]. Furthermore, keeping

in mind the anatomical meaning of certain landmarks (e.g., FME or FLE) as origin or insertion points of a ligament or tendon, multiple candidate landmark points can be considered equally correct, as no ligament or tendon is only attached to the bone by one single fiber. Thus, annotating landmarks is not a trivial task, as it is a simplification of a more complex anatomical reality. To warrant an optimal quality of the expert mean landmarks in the current study, the landmarks were annotated by three independent experts instead of one single observer. Secondly, the automated landmarking accuracy is related to the mesh resolution of the registered surfaces and the template mesh, as the candidate landmark positions are limited to the vertices of the registered surfaces. A higher resolution (more vertices) in the template mesh enhances accuracy but increases computing time for 3D surface registration. In addition, the resolution and slice thickness of the source imaging also play a role. Subtle ridges, such as the anterior border of the medial and lateral tibial plateau cartilage, might be more or less pronounced. Since our automated approach relies solely on the 3D models derived from thick-slice MRI scans, the TMCA and TLCA landmarks were more challenging to detect, resulting in larger AP TPM and AP TPL differences from the ground truth measurements. Finally, the surface registration algorithm is prone to the introduction of minor tangential translations of vertices over the 3D surface in nearly flat regions. This is reflected in landmark positioning errors, which do not contribute directly to larger measurement errors (as illustrated in Figure 9.8). A similar effect was also observed for the manual landmarking: while, e.g., the posterior points on the tibia plateau showed inter-observer differences up to 9.7 mm, the maximal anteroposterior size inter-observer differences were maximally only ca. 3.5 mm. It was indeed verified that the largest part of the TMCP and TLCP position differences is in the mediolateral and proximodistal coordinates.

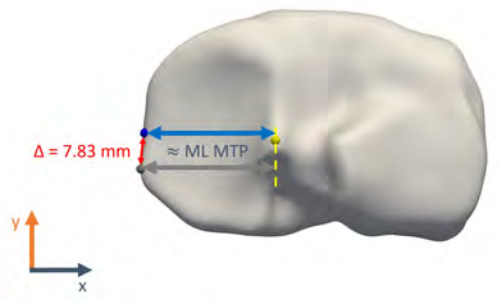


Figure 7.8: Exaggerated representation of a potential discrepancy between landmark position difference and corresponding measurements difference. Both blue and grey TMCM landmarks conform to the landmark definition, yet there is a 7.83 mm gap between them. In contrast, the calculated medial tibial plateau width only differs by 0.49 mm.

The main strength of our automated landmark tool is the extensive standardization of the complete process, thereby ruling out any intra- or inter-observer variation and making the time-consuming human input obsolete. The marginal operator time was reduced from up to 30 min per knee to zero after initialization of the landmarks of interest on the template shape. Combining a template-based method with some domain knowledge of the anatomy, a limited amount of training data were proven to be

sufficient. Our method does not require huge datasets for training, as is the case with deep learning approaches. While deep learning methods might be able to achieve higher accuracy, the training of these neural networks is slower and requires much more training data. Measuring distances and angles based on well-defined landmarks is probably the most straightforward and understandable method to analyze simple bone and cartilage shape variations. The clinical relevance of this type of bone shape analysis already lies within implant design and patient-specific pre-operative planning [23]. Using the 3D bone and cartilage models instead of the raw images is a potential limitation to evaluating the real-world clinical applicability of this method. Indeed, the focus of this study was on the landmarking process rather than the segmentation of the raw images. Using pre-operative MRI scans of the knee from routine diagnostic procedures for meniscus lesions from different clinical centers does ensure the robustness and generalization abilities of our method. Additionally, considering the data-driven nature of our method, a larger training dataset could be beneficial for automatic landmarking accuracy. However, similar validation studies report acceptable to excellent results based on similar sample sizes [20, 24, 25, 26]. The image quality of the clinical MRI scans (large slice thickness) could be considered a challenge for 3D bone and cartilage segmentation. Iterative verification and finetuning of the 3D models over three mutually perpendicular views were required to result in a manual segmentation accuracy of at least 1 mm [2]. Undoubtedly, ongoing advances in isotropic high-resolution MRI protocols [27] will provide higher-quality 3D models with more fine details (ridges and indents). This will facilitate both manual landmark annotation (a smaller region to focus on while searching the landmark) and automated landmarking (less prone to sliding surface errors during elastic deformation in the registration process). Future potential improvements include automating the segmentation process from raw images to reduce even further observer-related variability and human processing time. Furthermore, the surface registration algorithm could be accelerated by implementing multithread computing, but this was not the scope of this study. Surface registration remains the main time-consuming factor in the automated method and always entails a trade-off between accuracy and time consumption. A further potential enhancement is the introduction of Mean Value Coordinates [28], a generalization of barycentric coordinates. It overcomes the limitation of the registered mesh vertices being the only candidate landmark positions, allowing a lower template mesh resolution for a similar achievable landmarking accuracy.

7.5 Conclusion

In conclusion, considering the substantial variability among observers in the manual method, there is a clear need for an objective, operator-independent, and efficient approach to identifying anatomical landmarks. Our automated method demonstrated excellent accuracy and reliability for both landmark positioning and morphometric measurements. Moreover, this high level of automation will lead to a faster, scalable and human operator-independent morphometric analysis of the knee. Potential applications include optimized orthopedic implant designs, patient-specific treatment tailoring and large-scale morphometric risk factor analysis in different pathologies.

Bibliography

[1] Huang F, Harris S, Zhou T, Roby GB, Preston B, Rivière C. Which method for femoral component sizing when performing kinematic alignment TKA? An in silico study. Orthop Traumatol Surg Res. 2023;110:103769.

[2] Beeler S, Vlachopoulos L, Jud L, Sutter R, Fürnstahl P, Fucentese SF. Contralateral MRI scan can be used reliably for three-dimensional meniscus sizing—Retrospective analysis of 160 healthy menisci. Knee. 2019;26:954-61.

[3] Grammens J, Van Haver A, Danckaers F, Booth B, Sijbers J, Verdonk P. Small medial femoral condyle morphotype is associated with medial compartment degeneration and distinct morphological characteristics: A comparative pilot study. Knee Surg Sports Traumatol Arthrosc. 2021;29:1777-89.

[4] Van Haver A, De Roo K, De Beule M, Van Cauter S, Audenaert E, Claessens T, et al. Semi-automated landmark-based 3D analysis reveals new morphometric characteristics in the trochlear dysplastic femur. Knee Surg Sports Traumatol Arthrosc. 2014;22:2698-708.

[5] Peeters W, Van Haver A, Van den Wyngaert S, Verdonk P. A landmark-based 3D analysis reveals a narrower tibial plateau and patella in trochlear dysplastic knees. Knee Surg Sports Traumatol Arthrosc. 2020;28:2224-32.

[6] van der Merwe J, van den Heever DJ, Erasmus P. Variability, agreement and reliability of MRI knee landmarks. J Biomech. 2019;95:109309.

[7] Victor J, Van Doninck D, Labey L, Innocenti B, Parizel PM, Bellemans J. How precise can bony landmarks be determined on a CT scan of the knee? Knee. 2009;16:358-65.

[8] Fürmetz J, Sass J, Ferreira T, Jalali J, Kovacs L, Mück F, et al. Three-dimensional assessment of lower limb alignment: Accuracy and reliability. Knee. 2019;26:185-93.

[9] Bermejo E, Taniguchi K, Ogawa Y, Martos R, Valsecchi A, Mesejo P, et al. Automatic landmark annotation in 3D surface scans of skulls: Methodological proposal and reliability study. Comput Methods Programs Biomed. 2021;210:106380.

[10] Gupta A, Kharbanda OP, Sardana V, Balachandran R, Sardana HK. A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images. Int J Comput Assist Radiol Surg. 2015;10:1737-52.

[11] Danckaers F, Huysmans T, Lacko D, Ledda A, Verwulgen S, Van Dongen S, et al. Correspondence Preserving Elastic Surface Registration with Shape Model Prior. In: Proceedings of the 2014 22nd International Conference on Pattern Recognition; Stockholm, Sweden. 24–28 August 2014; 2014. p. 2143-8.

[12] Nguyen V, Alves Pereira LF, Liang Z, Mielke F, Van Houtte J, Sijbers J, et al. Automatic landmark detection and mapping for 2D/3D registration with BoneNet. Front Vet Sci. 2022;9:923449.

[13] Serafin M, Baldini B, Cabitza F, Carrafiello G, Baselli G, Del Fabbro M, et al. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: Systematic review and meta-analysis. Radiol Med. 2023;128:544-55.

[14] Valette S, Chassery JM, Prost R. Generic remeshing of 3D triangular meshes with metric-dependent discrete voronoi diagrams. IEEE Trans Vis Comput Graph. 2008;14:369-81.

[15] Schroeder W, Martin K, Lorensen B. The Visualization Toolkit. 4th ed. Kitware; Clifton Park, NY, USA; 2006.

[16] Gamer M, Lemon J, Fellows I, irr SP. Various Coefficients of Interrater Reliability and Agreement, R Package Version 0.84.1. 2024 February. Available from: `https://cran.r-project.org/web/packages/irr/irr.pdf`.

[17] Core R. Team . R: A Language and Environment for Statistical Computing (Version 4.2.1) R Foundation for Statistical Computing; Vienna, Austria; 2023.

[18] Koo TK, Li MYA. Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15:155-63.

[19] Martin Bland J, Altman Douglas G. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. Lancet. 1986;327:307-10.

[20] Kuiper RJA, Seevinck PR, Viergever MA, Weinans H, Sakkers RJB. Automatic Assessment of Lower-Limb Alignment from Computed Tomography. J Bone Jt Surg. 2023;105:700-12.

[21] Richard AH, Parks CL, Monson KL. Accuracy of standard craniometric measurements using multiple data formats. Forensic Sci Int. 2014;242:177-85.

[22] Smith AC, Boaks A. How "standardized" is standardized? A validation of postcranial landmark locations. J Forensic Sci. 2014;59:1457-65.

[23] MacLeod AR, Roberts SA, Gill HS, Mandalia VI. A simple formula to control posterior tibial slope during proximal tibial osteotomies. Clin Biomech. 2023;110:106125.

[24] Porto A, Rolfe S, Alpaca MAM. A fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures. Methods Ecol Evol. 2021;12:2129-44.

[25] Ridel A, Demeter F, Galland M, L'abbé E, Vandermeulen D, Oettlé A. Automatic landmarking as a convenient prerequisite for geometric morphometrics. Validation on cone beam computed tomography (CBCT)—Based shape analysis of the nasal complex. Forensic Sci Int. 2020;306:110095.

[26] Wilke F, Matthews H, Herrick N, Dopkins N, Claes P, Walsh S. Automated 3D Landmarking of the Skull: A Novel Approach for Craniofacial Analysis;.

[27] Van Dyck P, Smekens C, Roelant E, Vande Vyvere T, Snoeckx A, De Smet E. 3D CAIPIRINHA SPACE versus standard 2D TSE for routine knee MRI: A large-scale interchangeability study. Eur Radiol. 2022;32:6456-67.

[28] Seim H, Kainmueller D, Heller M, Zachow S, Hege HC. Automatic extraction of anatomical landmarks from medical image data: An evaluation of different methods. In: Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; Boston, MA, USA. 28 June–1 July 2009; 2009. p. 538-41.

Chapter 8

Statistical shape analysis: technical background

In this chapter, elementary concepts related to statistical shape analysis are provided. Starting from the meshes as generated by the marching cubes algorithm from the segmented image volumes, it covers the process of surface registration, Procrustes analysis to realign all cases in the dataset rigidly to the mean bone shape and finally the principal component analysis to extract the main modes of shape variation.

8.1 Introduction

Statistical Shape Modeling is a powerful tool used to capture, quantify, and analyze the variability of shapes in a population [1]. It has applications in various fields such as medical imaging [2], computer vision [3], and computational anatomy [4, 5, 6]. A statistical shape model (SSM) enables the identification of patterns and differences in shapes, facilitating tasks such as shape comparison, deformation analysis, and object recognition. The process involves multiple steps, from preprocessing 3D mesh data to dimensionality reduction techniques, such as Principal Component Analysis (PCA). Recently, deep learning approaches have also emerged, offering some promising innovations in shape analysis.

8.2 Dataset preparation

The first step in statistical shape modeling is preparing the 3D shapes (often represented as triangular meshes) for analysis (see also 5. Image segmentation and 3D meshing). These shapes can originate from medical imaging (e.g., MRI, CT scans) or other 3D scanning techniques. Mesh preprocessing ensures that the source and target shapes are clean and suitable for subsequent analysis.

- **Mesh Cleaning:** Raw meshes often contain noise, holes, or irregularities. Smoothing filters (e.g., Laplacian smoothing) and mesh repair algorithms are used to remove noise and fill gaps in the surface. [7, 8]

- **Resampling:** To ensure consistency, the number of vertices and faces in the meshes may be standardized across the dataset by resampling or remeshing techniques like isotropic remeshing, ensuring uniform density and distribution of points [9].
- **Alignment and Normalization:** Before surface registration, the source and target shapes can be aligned and normalized in terms of position, scale, and orientation. Techniques like Procrustes analysis, translation, and scaling are used to bring the meshes into a common frame of reference [10].

8.2.1 Procrustes analysis

The name of this statistical normalization technique originates from the bandit Procrustes in Greek mythology. He would offer travelers a place to stay, but once they lay down in his bed, he would either stretch their bodies or amputate limbs to ensure they fit the exact dimensions of the bed. This macabre legend became a metaphor for forced conformity, where something is altered to fit a rigid standard, sometimes at the expense of its natural form. In the mathematical analogy, the Procrustes analysis further used in this work "fits" one configuration of data to another, transforming it as needed but without distorting its intrinsic structure, by a combination of uniform scaling, translation and rotation.

For matters of completeness, an overview of several variants of Procrustes analysis is given:

- **Ordinary Procrustes Analysis (OPA)**:
  - This is the most basic form of Procrustes analysis. It aligns two sets of (surface) points by applying translation, uniform scaling, and rotation to one set of points to match the other. The goal is to minimize the sum of squared distances between corresponding points in the two sets.
  - Applications: OPA is often used in shape analysis, particularly in biology, where it helps to compare the shapes of organisms or anatomical features.
- **Generalized Procrustes Analysis (GPA)**:
  - GPA extends the concept to more than two sets of points. It simultaneously aligns multiple configurations (e.g., the shapes of multiple specimens) to a common average shape. This iterative process continues until the overall variance between the shapes is minimized.
  - Applications: GPA is widely used in anthropology, archaeology, and morphometrics to compare shapes across multiple individuals or groups, such as comparing skull shapes across different species.
- **Partial Procrustes Analysis (PPA)**:
  - PPA differs from OPA in that it does not allow for scaling. Only translation and rotation are applied to one set of points to match another. The emphasis here is on comparing shape similarity without altering the size.
  - Applications: Useful in cases where size is an important feature that should not be modified, such as comparing the growth patterns of different organisms or a comparison of knee morphology between distinct patient subgroups.

8.3 Surface registration

Surface registration is an essential step in statistical shape modeling, where the source and target shapes are aligned and correspondences between points on the surfaces are established. Accurate registration ensures that the shapes can be meaningfully compared (e.g. in terms of anatomical position), enabling further analysis of general shape variability. Multiple algorithms exist to perform surface registration, including iterative closest point (ICP) [11, 12], coherent point drift (CPD) [13] and normal distributions transform (NDT)[14, 15].
In this section, we focus on a variant of the iterative closest point algorithm, namely the elastic surface registration with or without a shape model prior, as described by Danckaers et al [16]. It builds further upon the work of Amberg et al. [17] for non-rigid ICP algorithm with translation vertices (N-ICP-T). It aims to deform the source surface to the target surface, while establishing a meaningful mapping between the vertices of the 3D surfaces, to preserve the anatomical correspondences, as illustrated in Figure 11.1.

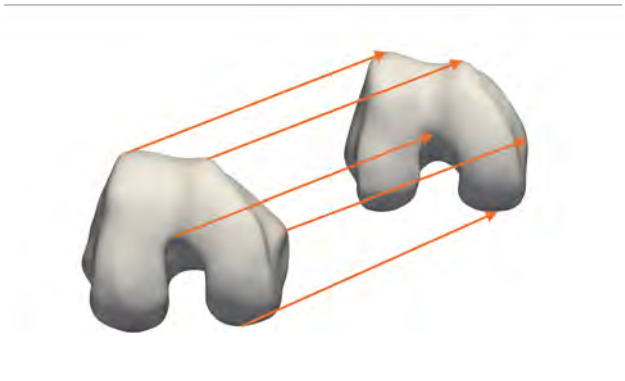


Figure 8.1: Visualisation of a subset of anatomical correspondences between source and target mesh.

8.3.1 Initial alignment

A first rigid alignment step helps to find corresponding points in a more efficient and accurate way. Multiple options exist to get the initial alignment roughly correct. Depending on the shape at hand, good results can be obtained by matching the centroids (points located at the mean coordinates of all surface points) and inertial axes (or principal axes, as obtained by principal component analysis on the 3D coordinates of the point cloud), eventually followed by an ICP rigid transformation step (with only a few iterations). The matching of principal axes might give suboptimal or even inferior results when the shape has two dimensions that are highly similar in terms of size. Applying only the centroid matching with iterative closest point rigid transformation might be a good alternative in that case. Usually, given a homogenous image (scan) dataset from a highly standardize data source, this approximate initial alignment is already obtained, when all instances were acquired in the same default reference coordinate system.

8.3.2 Corresponding point search

In this work, corresponding point search was performed by ray casting along the surface normals n_s from all vertices on the source mesh to the target mesh. Depending on the search strategy, either the intersection point self (not necessarily a vertex of the target mesh) or a set of closest points thereto are evaluated to match a number of criteria, including surface normal direction, surface curvature and distance between the source mesh vertex and candidate corresponding point.

Ray casting search strategy Surface normal direction is evaluated by calculating the dot product as defined in Eq. 11.1. If this dot product $a > a_{\min}$, then the intersection point is considered as a corresponding point. $a_{\min}$ gets more loose towards the end of the iterative process, going from 0.8 to 0.6, this is equivalent to a difference in normal direction from 36,9° to 53,1°, as calculated by $\arccos(a)$.

$$a = n_s \cdot n_t \tag{8.1}$$

Neighborhood search strategy In case a set of closest points to the intersection point is evaluated, a normalized surface curvature criterion is in place. First the curvature values (mean curvature is used, not Gaussian curvature) for all vertices on both shapes are normalized by subtracting the mean curvature (over all vertices from that mesh) and dividing by the standard deviation (again over all mesh vertices from that mesh) as defined in Eq. 11.2. Finally, a cost function defined as a weighted sum of the difference in normalized curvature and distance between source vertex and candidate corresponding point is used to withold the target vertex with the lowest cost as final corresponding point. An upper threshold to the absolute value of this cost function is in place and gets more strict towards the end of the iterative process, so there is no guarantuee that a corresponding point will be found for all vertices on the source mesh.

$$\kappa_{\text{norm}} = \frac{\kappa - \mu_\kappa}{\sigma_\kappa} \tag{8.2}$$

with μ_κ the mean curvature over the source or target surface, and σ_κ the standard deviation of the curvature over that surface.

8.3.3 Rigid transformation

Following the previously defined corresponding point search, an optimal affine transformation matrix is calculated. The distance between the source vertices with a corresponding point on the target surface and those corresponding points on the target surface is then minimized in a least-squares method. The resulting affine transformation matrix includes a combination of translations in the three dimensions, as well as rotations around the three axes of the reference coordinate system.

8.3.4 Elasticity modulated registration

After this rigid registration step, a new corresponding point search is performed. Assuming a more accurate rigid alignment, the corresponding point search is believed to result in turn in preciser and more correspondences to be established. These candidate corresponding points are then subject to an elastic deformation, where the vertices are allowed to translate separately, under the restriction of a stiffness parameter β. This stiffness parameter defines the strength of the connection between neighboring vertices, ensuring similar translations in neighboring vertices. It iteratively decreases towards a predefined lower bound in order to impose less constraint while evolving towards convergence. A diagonal weight matrix $W \in \mathbb{R}^{n \times n}$ keeps track of the detected correspondences. If no correspondences were detected for a source vertex, its respective weight is set to zero, to make it simply move along with its neighboring vertices. If a corresponding point was found the respective weight is set to be 1. Matrix $S \in \mathbb{R}^{n \times 3}$ and matrix $T \in \mathbb{R}^{n \times 3}$ represent the coordinates of the corresponding source and target vertices, respectively. The finally applied translation vectors, the matrix $X_T \in \mathbb{R}^{n \times 3}$, are then found by solving following linear system:

$$\begin{bmatrix} \beta C \\ W I_n \end{bmatrix} X_T = \begin{bmatrix} 0 \\ W(T - S) \end{bmatrix}, \tag{8.3}$$

with $C \in \mathbb{R}^{e \times n}$ the connectivity matrix or the incidence matrix of the target surface that indicates the start- and end vertex of each edge (connecting two vertices in the mesh). I_n is the $n \times n$ identity matrix. The rigid registration step and elastic registration steps are iteratively repeated together until the predefined number of registration steps is reached or until convergence, which is calculated by comparing the current distance between the source and target surface d_t and the previous distance between the source and target surface d_{t-1}. Convergence is then defined as follows when

$$\frac{|d_t - d_{t-1}|}{d_t} < 0.001.$$

8.3.5 Shape model prior modulated registration

The use of a shape model prior allows to let the source vertices move more freely within the constraints posed by the shape space as defined by the shape model prior. The final construction of a statistical shape model (SSM) is discussed in the following Section 11.4 Constructing a statistical shape model (SSM). To avoid inception-related misunderstandings: a shape model prior is not required to create a statistical shape model, it is only an alternative and more efficient way to expand a dataset of registered shapes or to redo the registration in order to achieve a higher registration accuracy (minimize distance between source and target surfaces, better geometric fit) for complex shapes.

Similarly to the rigid and elasticity modulated registration, the aim is to minimize the distance between the corresponding points, by fitting the shape model to the target surface in combination with a further elastic deformation. A predefined number of shape modes m is used, based the amount of explained variance or clinically meaningful modes of shape variation to incorporate. The idea behind this limitation is

to avoid the introduction of meaningless noise or corresponding point drift along the target surface, which is typically captured by the later modes, which correspond to a lower amount of total shape variance. At last, if a deformation of the source mesh along the largest and main modes of shape variation is already allowed, the remaining registration error can be further minimized by the elasticity modulated deformation. The iterative process now consists of three registration steps that are repeated: a rigid, a shape model prior modulated and an elasticity modulated registration step. A flowchart indicating the position of this shape model prior modulated registration step in the complete iterative process is shown in Figure 11.2.

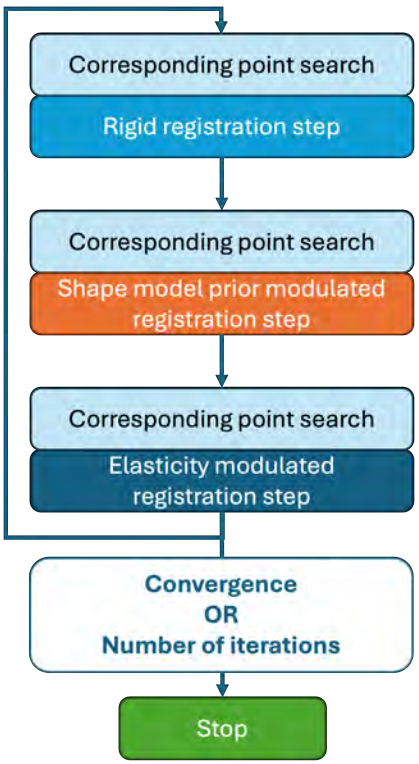


Figure 8.2: Schematic overview of the iterative process employing shape model prior modulated registration.

8.4 Constructing a statistical shape model (SSM)

A first requirement for constructing a statistical shape model (SSM) is that the meshes are in (anatomical) correspondence to each other and superimposed into the same reference coordinate system. This superposition of registered shapes is obtained by a partial procrustes analysis, which translates and rotates each shape instance to superimpose the mean shape.

8.4.1 Principal component analysis (PCA)

The model is then built by applying principal component analysis (PCA) on this dataset of corresponding point coordinates [1]. At its core, PCA is a method for reducing the dimensionality of a dataset while preserving the most important information. In the context of medical data, this might mean simplifying a complex set of measurements into a smaller number of meaningful components. The complex set of measurements in this case encompasses then the coordinates (x,y and z) of the m corresponding points over all shape instances. Indeed, the application of PCA in statistical shape analysis is only meaningful on a set of registered surfaces. That is, when the vertices of the surfaces are ordered in the same way from an anatomical point of view. Each shape is then represented by its set of corresponding points (or pseudo-landmarks), and after flattening of this $m \times 3$ matrix, as a one-dimensional shape vector (with length $3n$).

After registration, the mean shape is computed. The mean shape is simply the average of the corresponding coordinates of all the shapes in the dataset. This acts as the baseline shape for further analysis. If there are N shapes, each of them consisting out of m points, the mean shape $\bar{S} \in \mathbb{R}^{3m}$ is computed as:

$$\bar{S} = \frac{1}{N} \sum_{i=1}^{N} S_i \tag{8.4}$$

where $S_i \in \mathbb{R}^{3m}$ is the vectorized (flattened) representation of the i^{th} shape. In order to achieve a compact statistical shape model (see also 11.5: Model performance assessment) all shape instances are rigidly aligned to the mean shape prior to the construction of the covariance matrix.

Covariance matrix and singular value decomposition

First, a corresponding points matrix is constructed: $X \in \mathbb{R}^{N \times 3m}$ is given by

$$X = \begin{bmatrix} x_{0,0} & y_{0,0} & z_{0,0} & \dots & x_{0,m} & y_{0,m} & z_{0,m} \\ x_{1,0} & y_{1,0} & z_{1,0} & \dots & x_{1,m} & y_{1,m} & z_{1,m} \\ & \vdots & & \vdots & & \vdots & \\ x_{N,0} & y_{N,0} & z_{N,0} & \dots & x_{N,m} & y_{N,m} & z_{N,m} \end{bmatrix}, \tag{8.5}$$

with m the number of points and N the number of shapes. The i-th row of this corresponding points matrix represents the i-th registered shape of the dataset and is denoted by $x_i \in \mathbb{R}^{3m}$. In our practical example of knee bone shapes, the population of N knee bone shapes could then be seen as a point cloud of N points in an $3m$-dimensional space. This point cloud can be represented by a linear combination of $N - 1$ eigenmode vectors or principal components, where the first eigenmode refers to the direction of the largest variance in the analyzed dataset, the second eigenmode to the second largest variance perpendicular to the first, etc. Then, the **normalized** corresponding point matrix $\hat{X} \in \mathbb{R}^{N \times 3m}$ is obtained by subtracting the mean vectorized shape $\bar{S}$ from each row of the matrix, which finally will result in modeling of solely the surface deviations from the mean shape. Next, the normalized covariance matrix $C \in \mathbb{R}^{N \times N}$ is calculated:

$$C = \frac{1}{N-1} \hat{X} \hat{X}^T, \tag{8.6}$$

with eigenvalues $\Lambda \in \mathbb{R}^{N \times N}$ and eigenvectors $P \in \mathbb{R}^{3m \times N}$, as obtained from the SVD diagonalization of this covariance matrix C by

$$C = Q \Lambda Q^T, \tag{8.7}$$

with $\Lambda \in \mathbb{R}^{N \times N}$ the diagonal matrix holding the eigenvalues of the covariance matrix.

Given that $\hat{X}$ could also be diagonalized as follows:

$$\hat{X} = U \Sigma V^T, \tag{8.8}$$

it follows that:

$$C = \frac{1}{N-1} \hat{X} \hat{X}^T = \frac{1}{N-1} U \Sigma V^T \mathrm{V} \Sigma U^T = \mathrm{U} \frac{\Sigma^2}{N-1} U^T. \tag{8.9}$$

Taking into account Eq. 11.7, the singular values of $\hat{X}$ in Σ are closely related to the singular values λ_i on the diagonal of Λ via:

$$\lambda_i = \frac{{\sigma_i}^2}{N-1} \tag{8.10}$$

Furthermore, from Eq. 11.9, it follows that:

$$\hat{X} \hat{X}^T = U \Sigma^2 U^T, \tag{8.11}$$

and,

$$\hat{X} \hat{X}^T U = U \Sigma^2 \tag{8.12}$$

Therefore the left singular vector U as calculated by SVD on the matrix $\hat{X}$ is not only also an eigenvector of the covariance matrix C (Eq. 11.11), it equals the left singular vector Q of the covariance matrix C (Eq. 11.9). Finally, the principal components are in the columns of the matrix P, as calculated by:

$$P = \hat{X} \cdot Q \tag{8.13}$$

Any knee bone mesh $k \in \mathbb{R}^{3n}$, that lies in the shape space of the input knee bone meshes (e.g. either femur or tibia), can then be approximated by the sum of the average surface $\bar{S}$ and a linear combination of the principal components P as follows:

$$k = \bar{S} + P \cdot b, \tag{8.14}$$

where $b \in \mathbb{R}^{N-1}$ holds the knee bone shape parameters. To conclude, new knee bone shapes can be formed by varying in these shape model parameters b.

Dimensionality reduction

By keeping only the first few eigenvectors (those corresponding to the largest eigenvalues), we can reduce the dimensionality of the shape space. For example, if 10 principal components explain 95% of the variance, the shapes can be represented in a 10-dimensional space instead of a $3m$-dimensional space with m the number of vertices.

The new low-dimensional approximation $S_i' \in$ of each shape S_i is then given by projecting the shape onto the l first principal components:

$$S_i' = \bar{S} + \sum_{k=1}^{l} b_{ik} P_k \tag{8.15}$$

where b_{ik} are the shape-specific weights (scores) for the i^{th} shape along the k^{th} principal component.

8.4.2 PPSA

For matter of completeness, the concept of PPSA is shortly introduced here, the reader is referred to the work of Duquesne et al. for more technical details [18]. Traditional shape analysis techniques employing PCA look for linear modes of variation in these shapes. Essentially, it assumes that the changes in the landmarks follow straight-line paths when one shape is deformed into another. PCA reduces the dimensionality of the data by identifying these linear combinations of features that explain the most variance in the shape. However, in many cases, the way shapes vary is not linear. For instance, the biological growth of an organism, or dynamic movement patterns in human gait, may result in more complex, curving deformations that linear methods cannot fully capture.

This is where Principal Polynomial Shape Analysis becomes useful. PPSA extends the linear framework of PCA by introducing polynomial terms to describe more complex, curvilinear relationships between the landmarks. Instead of assuming that landmarks move in straight lines as a shape deforms, PPSA allows for the possibility that these landmarks move along curved paths, following a polynomial function.

The core idea behind PPSA is to capture higher-order relationships between the landmarks. A polynomial is a mathematical function that can model curves and more complex patterns of movement. For example, a quadratic polynomial (which includes squared terms) can model shapes that expand or contract in a way that forms a parabolic curve. A cubic polynomial (which includes terms raised to the third power) can capture even more complex deformations, such as twisting or bending motions.

Despite its power, there are some challenges associated with Principal Polynomial Shape Analysis. One of the main challenges is selecting the right degree for the polynomial terms. A low-degree polynomial might not capture enough of the variation, while a high-degree polynomial might overfit the data, meaning it could model random noise rather than true underlying patterns. Therefore, careful selection of the polynomial degree is critical to ensure the analysis is accurate and meaningful.

Another potential challenge is the interpretability. While the results of PPSA can reveal more complex modes of variation, the inclusion of polynomial terms can make it harder to interpret those modes in simple, intuitive ways. In PCA, the principal components correspond to easily understood linear transformations of the data. In PPSA, the modes of variation involve more complex, curvilinear transformations, which might require more effort to explain in terms of the underlying biological or geometric processes.

As the analyzed knee bone shapes in this work were only analyzed mainly in their own reference coordinate system, and no actual kinematic analysis was performed, just

simple PCA was used to extract the most meaningful modes of shape variation. Furthermore, there were no indications of PCA failing to capture the shape variations in terms of compactness (explained variance), generalizability or specificity.

8.5 Model performance assessment

To evaluate the effectiveness of an SSM, the following key metrics are typically employed: compactness, generalization to unseen shapes and specificity. Naturally, all of these metrics can be expressed in function of the number of modes of shape variation taken into account.

Model compactness refers to the ability of the model to represent the underlying variability of the shape with the fewest number of parameters. A compact model can accurately represent the variation of shapes in the dataset using a reduced number of modes or principal components. This is a result of the dimensionality reduction performed by PCA, which captures the most significant variations in shape with fewer dimensions. A more compact model is desirable because it reduces the computational burden and avoids overfitting by preventing the model from becoming too complex or sensitive to noise in the training data [1]. However, compactness needs to be balanced with the ability of the model to capture sufficient variation in the dataset.

Next, **model generalizability** is another key criterion that measures how well the statistical shape model can represent new, unseen instances of shapes from the same class. In other words, a generalizable model should accurately capture shapes that were not part of the training set, reflecting the true variability of the population [19]. A well-constructed model must generalize beyond the training data, as it ensures applicability to broader, real-world scenarios. This can be quantitatively assessed by testing the model on a set of unseen data and observing how well it represents those new instances. Generalizability is heavily influenced by the training set: a model trained on insufficient or unrepresentative data may fail to generalize, leading to poor performance in practical applications.

Finally, **model specificity** complements generalizability by focusing on how well the SSM restricts itself to producing valid shapes from the learned class [19]. In an ideal scenario, the model should not generate shapes that deviate significantly from the true anatomical or object shape variability. A highly specific model will only represent shapes that are plausible according to the dataset it was trained on, effectively ruling out unlikely or biologically implausible shapes. This property can be evaluated by generating random instances of shapes from the model and measuring how similar these are to real shapes from the target population [20]. A lack of specificity indicates that the model is overly flexible and might produce unrealistic shapes, which can undermine the utility of the model in clinical or engineering applications.

While these three discussed performance metrics are a good first check of the model's ability to describe the complete shape space by a confined number of principal components, these metrics might be misleading when assessing the quality of the underlying correspondences in some cases [20, 21]. Optimization of the registration parameters allows to minimize the surface distance between the deformed source mesh

and the target mesh. Therefore, registration errors along the surface normal direction are usually neglectable. However, correspondence errors in the tangential direction across the surface, cannot be ruled out. As it is a truly impossible job to manually define an anatomical correspondence ground truth for thousands of vertices, this validation step could be reduced to evaluate a confined number of anatomical landmarks. Having a similar vertex density (as imposed by the initial isotropic remeshing step and maintained by the stiffness parameter of the elasticity modulated non-rigid deformation) across all surface regions and instances, it is an acceptable hypothesis that the potential tangential correspondence error for all points is in the same order of magnitude as for this confined number of anatomical landmarks.

8.6 Deep learning: the future?

Deep learning offers substantial advancements over traditional methods, especially in its ability to model complex shapes and handle large, high-dimensional data. Traditional statistical shape analysis methods, such as Procrustes analysis or principal component analysis (PCA) on (pseudo-)landmarks, rely on linear assumptions and often require careful pre-processing steps like landmark placement, or surface registration. These approaches are effective for simple or well-understood shape distributions but struggle to capture intricate, non-linear variations in shape that are common in real-world objects, particularly when dealing with complex biological structures, 3D surfaces, or medical imaging data.

Deep learning methods, especially convolutional neural networks (CNNs), autoencoders, and generative models, have demonstrated remarkable success in overcoming these limitations by learning to represent and generate shapes directly from point cloud data without the need for explicit landmarking or surface registration [22]. They can automatically extract and encode high-level features, capturing complex, non-linear shape variations that traditional methods would overlook. This is particularly useful for high-dimensional data such as 3D meshes or volumetric data, where traditional statistical methods would face challenges in terms of computational cost and scalability.

A significant development in this area is **geometric deep learning**, which extends deep learning techniques to non-Euclidean domains such as graphs, manifolds, and point clouds—structures that are central to many shape analysis tasks. Traditional deep learning methods like CNNs are inherently designed for Euclidean grids, such as pixel-based images. However, shapes, especially those represented by 3D surfaces or anatomical structures, often lie on curved manifolds or are represented as graphs (e.g., skeletal structures or triangular meshes). Geometric deep learning employs specialized architectures, such as graph convolutional networks (GCNs) and spectral methods, to operate directly on these non-Euclidean structures. For instance, GCNs adapt the convolutional operation to graph-structured data by aggregating information from a node's neighborhood, thereby allowing the network to capture local geometric properties[23]. This has led to improved performance in tasks like shape correspondence and classification.

Autoencoders and **variational autoencoders (VAEs)** are particularly effective in

statistical shape analysis for learning compact, latent representations of shapes. These neural network-based models automatically learn to compress shape data into a low-dimensional latent space, capturing complex shape deformations or variations in a manner that is inherently non-linear [24]. This contrasts with traditional PCA-based methods, which can only capture linear modes of variation. This means they can only represent a subject shape as a weighted sum of modes of shape variations. VAEs extend autoencoders by introducing a probabilistic framework that enables the generation of new shapes by sampling from the latent space [25], thus supporting tasks such as shape generation, interpolation, and morphing [26]. These generative models furthermore allow a more flexible and scalable approach to shape modeling, capable of learning from large, unstructured datasets even without the need for explicit shape priors [27, 28].

The computational cost, in combination with the explainability of the models and final results are the main challenges in these approaches. However, the technique holds promise for the future. Firstly, deep learning methods can handle large-scale datasets, learning directly from raw data and requiring minimal manual feature engineering or pre-processing. Furthermore, they are highly flexible and can model complex, non-linear shape variations, which are often present in real-world shapes, making them robust in applications such as medical imaging or biological structure analysis.

Bibliography

[1] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models-Their Training and Application. Comput Vis Image Underst. 1995 1;61:38-59.

[2] Schmid J, Assassi L, Chênes C. A novel image augmentation based on statistical shape and intensity models: application to the segmentation of hip bones from CT images. European Radiology Experimental. 2023 8;7:39.

[3] Audenaert EA, Houcke JV, Almeida DF, Paelinck L, Peiffer M, Steenackers G, et al. Cascaded statistical shape model based segmentation of the full lower limb in CT. Computer Methods in Biomechanics and Biomedical Engineering. 2019 4;22:644-57.

[4] Meynen A, Matthews H, Nauwelaers N, Claes P, Mulier M, Scheys L. Accurate reconstructions of pelvic defects and discontinuities using statistical shape models. Computer methods in biomechanics and biomedical engineering. 2020 10;23:1026-33.

[5] Angelis SD, Henckel J, Bergiers S, Hothi H, Laura AD, Hart A. Statistical shape modeling of the large acetabular defect in hip revision surgery. Journal of Orthopaedic Research. 2023 9;41:2016-25.

[6] Eggermont F, Mathijssen E, Bakker M, Tanck E. Using a statistical shape model to estimate the knee landmarks for aligning femurs for femoral finite element models. Computer methods and programs in biomedicine. 2024 10;255:108324.

[7] Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. MeshLab: an Open-Source Mesh Processing Tool. In: Scarano V, Chiara RD, Erra U, editors. Eurographics Italian Chapter Conference. The Eurographics Association; 2008. .

[8] Schroeder W, Martin K, Lorensen B. The Visualization Toolkit. 4th ed.; 2006.

[9] Hoppe H, DeRose T, Duchamp T, McDonald J, Stuetzle W. Mesh optimization. In: Proceedings of the 20th annual conference on Computer graphics and interactive techniques. ACM; 1993. p. 19-26.

[10] Gower JC. Generalized procrustes analysis. Psychometrika. 1975 3;40:33-51.

[11] Besl PJ, McKay ND. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1992;14(2):239-56.

[12] Chen Y, Medioni G. Object modelling by registration of multiple range images. Image and Vision Computing. 1992 4;10:145-55.

[13] Myronenko A, Song X. Point Set Registration: Coherent Point Drift. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010 Dec;32(12):2262–2275.

[14] Biber P, Strasser W. The normal distributions transform: a new approach to laser scan matching. In: Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453). vol. 3; 2003. p. 2743-8 vol.3.

[15] Magnusson M, Andreasson H, Nuchter A, Lilienthal AJ. Appearance-based loop detection from 3D laser data using the normal distributions transform. In: 2009 IEEE International Conference on Robotics and Automation; 2009. p. 23-8.

[16] Danckaers F, Huysmans T, Lacko D, Ledda A, Verwulgent S, Dongen SV, et al. Correspondence Preserving Elastic Surface Registration with Shape Model Prior. In: 2014 22nd International Conference on Pattern Recognition. IEEE; 2014. p. 2143-8.

[17] Amberg B, Romdhani S, Vetter T. Optimal Step Nonrigid ICP Algorithms for Surface Registration. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2007. p. 1-8.

[18] Duquesne K, Nauwelaers N, Claes P, Audenaert E. Principal polynomial shape analysis: A non-linear tool for statistical shape modeling. Computer Methods and Programs in Biomedicine. 2022;220:106812.

[19] Goparaju A, Csecs I, Morris A, Kholmovski E, Marrouche N, Whitaker R, et al. On the Evaluation and Validation of Off-the-Shelf Statistical Shape Modeling Tools: A Clinical Application. In: Reuter M, Wachinger C, Lombaert H, Paniagua B, Lüthi M, Egger B, editors. Shape in Medical Imaging. Springer International Publishing; 2018. p. 14-27.

[20] Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis. 2009 8;13:543-63.

[21] Ericsson A, Karlsson J. Benchmarking of algorithms for automatic correspondence localisation; 2006. British Machine Vision Conference, 2006, BMVC 2006 ; Conference date: 04-09-2006 Through 07-09-2006.

[22] Croquet B, Matthews H, Mertens J, Fan Y, Nauwelaers N, Mahdi S, et al. Automated landmarking for palatal shape analysis using geometric deep learning. Orthodontics & craniofacial research. 2021 12;24 Suppl 2:144-52.

[23] Qi CR, Su H, Mo K, Guibas LJ. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation; 2017. Available from: https://arxiv.org/abs/1612.00593.

[24] Ehrhardt J, Wilms M. In: Autoencoders and variational autoencoders in medical image analysis. Elsevier; 2022. p. 129-62.

[25] Kingma DP, Welling M. Auto-Encoding Variational Bayes; 2022. Available from: https://arxiv.org/abs/1312.6114.

[26] Gutiérrez-Becker B, Sarasua I, Wachinger C. Discriminative and generative models for anatomical shape analysis on point clouds with deep neural networks. Medical Image Analysis. 2021 1;67:101852.

[27] Nauwelaers N, Matthews H, Fan Y, Croquet B, Hoskens H, Mahdi S, et al. Exploring palatal and dental shape variation with 3D shape analysis and geometric deep learning. Orthodontics & craniofacial research. 2021 12;24 Suppl 2:134-43.

[28] Yuan M, Hoskens H, Goovaerts S, Herrick N, Shriver MD, Walsh S, et al. Hybrid autoencoder with orthogonal latent space for robust population structure inference. Scientific reports. 2023 2;13:2612.

Chapter 9

Elementary concepts of machine learning

This chapter contains a very general introduction to some basic concepts in machine learning. It covers the dataset cleaning and preprocessing, feature engineering, data flow trough the model during training versus testing, hyperparameter optimization to avoid overfitting and various (cross-validation) evaluation strategies.

9.1 Introduction

Machine learning (ML) is a subfield of artificial intelligence (AI) that empowers computers to learn from data without explicit (rule) programming (Figure 13.1). By identifying patterns and relationships within datasets, ML algorithms can make predictions or decisions on new, unseen data. This capability has revolutionized various industries, from healthcare to finance.

There are two primary approaches to ML [2]:

- Supervised Learning: In supervised learning, the algorithm is trained on a labeled dataset, where each data point is paired with a corresponding target value or label. The algorithm learns to map input data to output labels by identifying patterns and relationships between the features and the target values. Common supervised learning algorithms include linear regression, logistic regression, decision trees, random forests, and support vector machines (SVMs). For instance, a linear regression model can be used to predict house prices based on features like square footage, number of bedrooms, and location.
- Unsupervised Learning: In unsupervised learning, the algorithm is trained on an unlabeled dataset, where the data points do not have corresponding target values. The algorithm learns to identify patterns and structures within the data itself, without any external guidance. A commonly used family of unsupervised learning algorithms is clustering. K-means clustering, for example, can be used to group customers into segments based on their purchasing behavior.

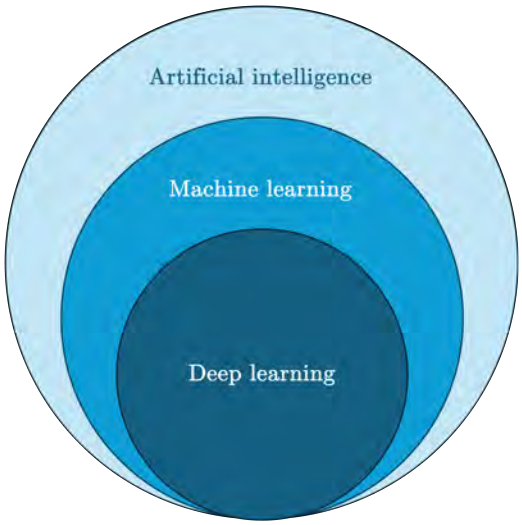


Figure 9.1: Schematic overview of how machine learning relates to artificial intelligence (AI) and deep learning. Adapted from [1].

Deep Learning: a subset of ML techniques

As earlier explained in Section "6.2.1.Deep learning", deep learning is a subset of ML (Figure 13.1) that utilizes neural networks with multiple layers to learn complex patterns from data. These neural networks are inspired by the structure and function of the human brain, with interconnected nodes that process information in a hierarchical manner [3]. Deep learning models can learn from large amounts of data and extract high-level features automatically, making them particularly effective for tasks such as image recognition, natural language processing, and speech recognition. They are typically well-suited to process vast amounts of information as input, both in terms of number of features as in terms of number of observations or cases.

Some considerations to take into account when choosing between traditional ML or deep learning techniques [2]:

- Deep learning models are often more powerful than traditional ML models, especially when dealing with complex patterns such as biomedical images and natural language processing.
- Deep learning models can automatically learn high-level features from data, reducing the need for manual feature engineering.
- Deep learning models are computationally more expensive to train and require generally larger amounts of data than traditional ML approaches.
- Traditional ML models are often more interpretable than deep learning models, making it easier to understand how they make decisions. After all, the automatically learned high-level features by deep learning models are not required to a have physical meaning.
- Traditional ML models usually require less computing power, thereby imposing less requirements on the hardware. Depending on the used hardware and dataset

size, both model training and inference tend to be faster with traditional machine learning approaches.

In summary, both traditional ML and deep learning have their respective strengths and weaknesses, and the choice of approach depends on the specific task at hand and the available resources (hardware and training dataset characteristics). While deep learning is often hyped and considered more performant these days, traditional ML algorithms remain valuable for many applications, given their easier to fulfill hardware requirements, their explainability and their lower tendency to overfit to the training dataset.

9.2 Dataset preprocessing

Raw data, in a healthcare context typically originating from the patient file, is often noisy, incomplete, inconsistent, or may even contain irrelevant features for the problem at hand. Thus, before it can be used to train a predictive model, dataset preprocessing is an essential step to clean, transform and prepare the dataset to extract its full potential [4].

While the specific steps involved in preprocessing may vary depending on the nature of the data and the machine learning task, there are some commonly adopted practices [5]. These steps ensure the data's quality and help in optimizing the model's performance.

9.2.1 Manual data checks and the role of domain expertise

While automated preprocessing techniques can manage much of the data preparation, manual verification and exploratory data analysis remains crucial [6]. This process involves the quality assessment of the data, examining outliers, data distributions, and potential anomalies that could adversely affect model training. For example, an unusually high or low value in a physiological measurement (e.g., cartilage thickness) could reflect a genuine clinical abnormality or be the result of a data entry error. Domain expertise is extremely helpful in differentiating between significant outliers and mistakes in the data. In surgery outcome prediction, clinicians play a vital role in validating the data, ensuring that the variables incorporated into the model are clinically relevant and that preprocessing maintains the integrity of the medical information.

Visualizing the dataset through tools such as histograms, box plots, and correlation matrices is a highly effective method for understanding the data. It helps to reveal patterns, identify outliers, skewed distributions and highlight relationships between features, such as redundancies. These valuable insights allow for more informed decisions during feature engineering and serve as one of the foundations for designing an efficient predictive model.

9.2.2 Data cleaning

In addition to manual dataset review, automated data cleaning methods are in place. Data cleaning is the process of identifying and rectifying (human) errors in the dataset [7]. It also includes handling missing values, next to correcting erroneous data, and dealing with noisy data. Methods for dealing with missing values include:

- remove instances or features with missing data if their absence is significant and likely to distort the analysis
- imputation techniques, such as filling missing values with the mean or median. Also more advanced methods such as k-nearest neighbors can be applied.

For noisy or erroneous data, methods like smoothing, binning or outlier detection algorithms can help ensure that the data does not adversely affect the learning process. The specific decision to clean or discard noisy data is context-specific, depending on the task at hand.

9.2.3 Feature scaling

Many machine learning algorithms, especially those based on distance metrics (e.g., k-nearest neighbors, support vector machines,...) are sensitive to the range and scale of the features [8, 9]. If the features have different ranges, the model may prioritize certain features, distorting the learned relationships. Therefore, scaling features to a standard range (e.g., between 0 and 1, or standardizing to z-scores with zero mean and unit variance) is essential to ensure that no feature dominates the others due to its scale. Furthermore, this might also help in reaching convergence faster, even in less sensitive algorithms such as logistic regression.

9.2.4 Feature encoding

Analoguous to conventional statistical methods, categorical features need to be transformed into a numerical format to become a suitable input for machine learning algorithms [10]. Two common techniques for encoding are:

- one-hot encoding: where each category is represented by a binary vector, ensuring no ordinal relationships are assumed
- label encoding: where each category is assigned a unique integer. This method is suitable when the categorical variable is ordinal, meaning that the values have an inherent order.

9.2.5 Feature selection

Feature selection involves identifying the most relevant features for the learning task [11]. Irrelevant or redundant features may increase the dimensionality of the dataset,

leading to slower training times, increased model complexity, and overfitting [12]. Non-informative features contribute mostly to the noise in the dataset and should be therefore be avoided from the input of machine learning algorithms. Dimensionality reduction techniques, such as principal component analysis (PCA) or linear discriminant analysis (LDA), can be employed to reduce the feature space while preserving as much variance or discriminative power as possible. Alternatively, statistical methods like correlation analysis or mutual information can help to identify which features are redundant.

9.2.6 Data transformation and augmentation

In many cases, data needs to be transformed to ensure it fits the assumptions of the learning algorithm [8, 5]. For instance, certain algorithms assume normally distributed input, making log transformations or power transformations useful for non-Gaussian features. Additionally, in scenarios where the dataset is imbalanced, such as in rare event detection, data augmentation techniques like oversampling the minority class or undersampling the majority class can be employed to balance the dataset. This step ensures that the model learns equally from all classes, preventing bias toward more frequent outcomes.

9.3 Data splits: train, validation and test sets

After the data preprocessing, it must be split into three subsets [2]:

- Training set: used to train the model, this is usually the largest portion of the data and is what the model learns from. The model uses this subset to learn patterns and relationships between input features and the target variable.
- Validation set: a smaller portion of the data, used during the model training phase for tuning hyperparameters (see also 13.4 and preventing overfitting. The model does not ”see” the validation set during training, but only used it to measure performance after each training iteration for a specific set of hyperparameters, defining e.g. the amount of regularization. Based on these performance results, the optimal set of hyperparameters is chosen.
- Test set: the final subset, unseen by the model during training and validation (and hyperparameter selection), is used to evaluate the model’s generalization performance. This set simulates new, real-world, unseen data and gives an unbiased estimate of how well the model will perform on new data.

It is of crucial importance these specific datasets remain strictly separated and are only used for what they are intended to. When other/more data than the training data is used to create the model (typical example is during preprocessing, to scale features to a certain range based on the full dataset instead of only the training dataset) this is called data leakage. It causes overly optimistic models, that might become completely invalid in production environments on unseen data. It is therefore common practice to use

strictly separated data pipelines during model training, intermediate validation (and hyperparameter optimization) and final model evaluation.

9.4 Hyperparameter optimization

Hyperparameters are the settings that control the learning process but are not learned by the model during training [13]. These include things like number of principal components to retain, number of hidden layers in a neural network, and the type of penalty terms in regularization. Some frequently used techniques to find the best combination of hyperparameters include [14]:

- Grid Search: One approach to finding the best hyperparameters is to manually define a range of possible values for each hyperparameter and search through all combinations. This is known as grid search, which is exhaustive but can be computationally expensive for large parameter spaces (Figure 13.2).
- Random Search: Random search selects random combinations of hyperparameters within a defined space, often leading to better results in less time compared to grid search. (Figure 13.2).
- Bayesian Optimization: Without going too much into detail, this is a more advanced technique that uses probabilistic models to predict the performance of hyperparameter combinations and iteratively updates its search based on previous results.

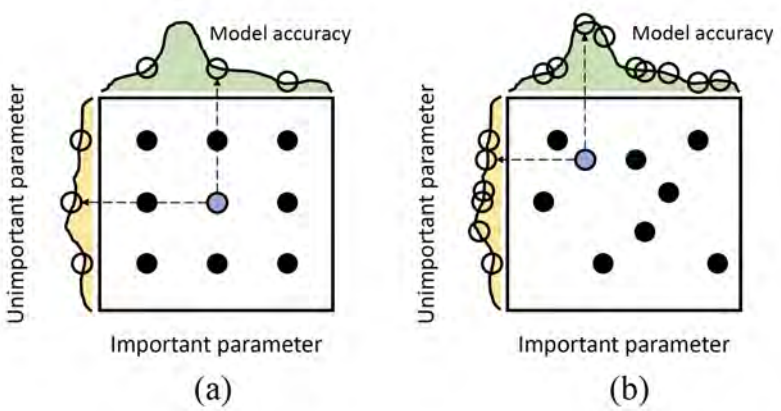


Figure 9.2: Comparison between **(a)** the grid search and **(b)** the random search strategy for sampling the hyperparameter landscape. While the random search allows less control of the multiple hyperparameter combinations to probe, it is clear from the figure that it is presumed to probe more distinct values for each hyperparameter individually. The curves (yellow and green) indicate the model's performance in function of the assessed hyperparameter. Figure from [15].

9.5 Model training

Model training is the heart of the machine learning process, where the algorithm learns from data by adjusting its internal parameters. During this phase, the model discovers patterns in the training subset and learns to make predictions based on input features. A loss function is used to calculate the error between the model prediction and the actual target for cases in the validation subset. Next, the optimization algorithm (e.g. stochastic gradient descent, ADAM,...) seeks new internal parameters of the model, based on the result of this loss function result. This is a step which is often referred to as backpropagation. It ensures that each internal parameter of the model is updated correctly, based on how much it contributes to the error at the output (and the chosen learning rate).
During model training, the algorithm typically passes multiple times over the entire training dataset, often referred to as epochs. The training process continues for several epochs until the model converges, meaning that the loss function reaches a minimum and further training does not significantly improve performance.

9.5.1 Avoiding overfitting

While training, the model can become too good at fitting the training data, capturing noise and spurious patterns that do not generalize well to unseen data [16, 17]. This is known as overfitting. Techniques to avoid overfitting during training include:

- Regularization: Adding a penalty to the loss function that discourages overly complex models.
- Early Stopping: Monitoring the model's performance on the validation set and stopping training when performance starts to degrade, indicating overfitting.
- Dropout: A technique used in neural networks where randomly selected neurons are ”dropped” during training to prevent co-adaptation and promote generalization.

9.6 Model evaluation

Model evaluation is a critical step in the machine learning pipeline that assesses how well a model performs on unseen data. It involves analyzing the model's predictive accuracy, robustness, and generalization ability. Effective evaluation helps ensure that the model not only performs well on training data but also generalizes effectively to new, real-world data. This process involves using a variety of metrics and strategies tailored to the specific type of problem being addressed—whether classification, regression, or another task.

9.6.1 Evaluation Metrics

The choice of evaluation metrics depends on the nature of the problem. Here's a comprehensive overview of metrics commonly used for classification tasks [18]:

1. **Accuracy:**
   - Measures the proportion of correctly predicted instances out of the total instances.
   - Formula:
   $$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$
   - Suitable for balanced datasets but can be misleading in cases of class imbalance.
2. **Precision and Recall:**
   - **Precision:** The proportion of true positive predictions out of all positive predictions made by the model.
   $$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$
   - **Recall:** The proportion of true positive predictions out of all actual positive instances.
   $$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$
   - Precision is crucial when the cost of false positives is high, while recall is important when missing a positive instance has severe consequences.
3. **F1-Score**:
   - The harmonic mean of precision and recall, providing a single metric that balances both aspects.
   - Formula:
   $$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
   - Useful when the classes are imbalanced and there is a need to balance precision and recall.
4. **ROC-AUC (Receiver Operating Characteristic - Area Under Curve):**
   - Measures the model's ability to discriminate between positive and negative classes across different thresholds.
   - The ROC curve plots the true positive rate (recall) against the false positive rate.
   - AUC represents the area under the ROC curve and ranges from 0 to 1, with higher values indicating better model performance.
5. **Confusion Matrix:**

- A table that summarizes the performance of a classification model by displaying the counts of true positives, true negatives, false positives, and false negatives.
- It provides a comprehensive view of model performance and helps in calculating various metrics like precision, recall, and F1-score.

9.6.2 Evaluation Strategies

Effective model evaluation involves more than just calculating metrics. Just as important as the metrics themselves, is the subset of data where they are calculated [19, 20]. Various strategies are available to ensure the model's performance is reliable and generalizes well [16]:

1. **Train-Test Split:**
   The simplest evaluation strategy involves splitting the dataset into training and testing subsets. The model is trained on the training set and evaluated on the test set to assess its performance on unseen data.
2. **Bootstrapping:**
   Involves repeatedly sampling from the dataset with replacement to estimate the variability of the model's performance. This technique is useful for understanding the stability and reliability of the model. It is a method to compute confidence intervals (CI) for a models performance metrics.
3. **Cross-Validation:**
   - **K-Fold Cross-Validation:** The dataset is divided into k subsets (folds). The model is trained k times, each time using $k - 1$ folds for training and the remaining fold for testing. The average performance across all folds provides a robust estimate of model performance.
   - **Leave-One-Out Cross-Validation (LOOCV):** A special case of k-fold cross-validation where k equals the number of instances in the dataset. Each instance is used as a test set once, and the remaining instances are used for training. It is typically used when the training dataset is small and collecting additional data is expensive. The downside of this method is the computational cost, as the model needs to be trained and evaluated on a lot of distinct subsets.
   - **Stratified Cross-Validation:**
     An extension of k-fold cross-validation that ensures each fold has a similar proportion of each class, particularly useful for imbalanced datasets.
4. **Nested Cross-Validation:**
   In order to get a grasp on the model's generalizability to unseen data when the total dataset is rather small, nested cross-validation makes full use of all available data. Nested k-fold cross-validation employs two loops: an outer loop consisting of k data folds and an inner loop of j folds. The inner loop is then used for hyperparameter optimization (see 13.4: Hyperparameter optimization), model training and model selection, while the outer loop only serves to assess the generalizability of the final model to unseen data (Figure 13.3).

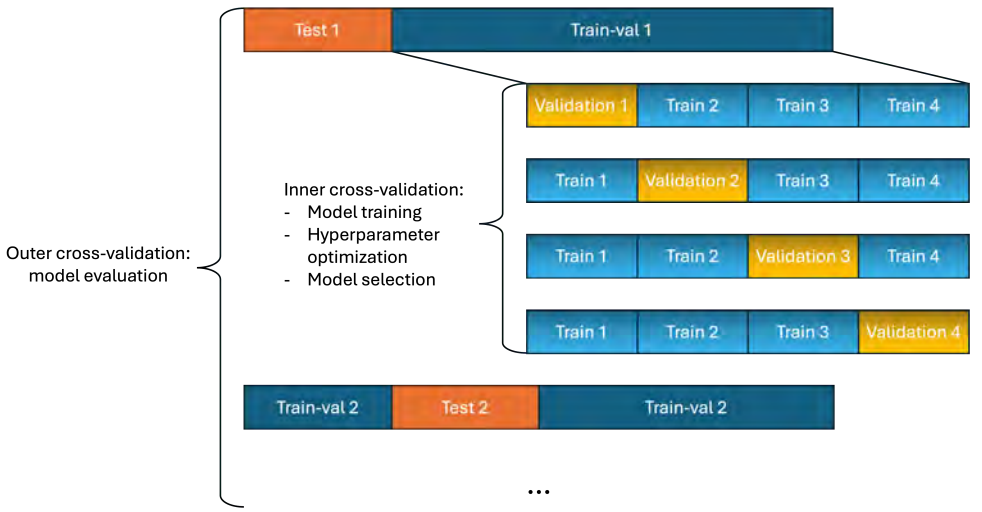


Figure 9.3: Schematic overview nested cross-validation with 4 outer and 4 inner folds: the validation subsets (yellow, inner loop) are used for hyperparameter optimization and model selection, while the actual test subsets (orange, outer loop) are only used for the assessment of the final model's generalizability to unseen data.

9.7 Conclusion

In machine learning, the key to building a successful model lies in properly splitting the data, training the model while carefully tuning its hyperparameters, and finally, evaluating its performance on unseen data. By using robust techniques for model evaluation, it can be assured the final model generalizes well and delivers accurate predictions in real-world scenarios.

Bibliography

[1] Shahzad U. A comparative analysis of artificial neural network and support vector machine for online transient stability prediction considering uncertainties. Australian Journal of Electrical and Electronics Engineering. 2022 4;19:101-16.

[2] Alpaydin E. Introduction to machine learning. MIT press; 2020.

[3] Goodfellow IJ, Bengio Y, Courville A. Deep Learning. Cambridge, MA, USA: MIT Press; 2016.

[4] Pfob A, Lu SC, Sidey-Gibbons C. Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. BMC Medical Research Methodology. 2022 11;22:282.

[5] Popov A. 1 - Feature engineering methods. In: Pal K, Ari S, Bit A, Bhattacharyya S, editors. Advanced Methods in Biomedical Signal Processing and Analysis. Academic Press; 2023. p. 1-29.

[6] Cox V, Cox V. Exploratory Data Analysis: What Data Do I Have? Translating Statistics to Make Decisions: A Guide for the Non-Statistician. 2017:47-74.

[7] Rahmani AM, Yousefpoor E, Yousefpoor MS, Mehmood Z, Haider A, Hosseinzadeh M, et al. Machine Learning (ML) in Medicine: Review, Applications, and Challenges. Mathematics. 2021 11;9:2970.

[8] scikit-learn developers. 6.3. Preprocessing data; 2024. Available from: `https://scikit-learn.org/stable/modules/preprocessing.html`.

[9] Patro S. Normalization: A preprocessing stage. arXiv preprint arXiv:150306462. 2015.

[10] Potdar K, Pardawala TS, Pai CD. A comparative study of categorical variable encoding techniques for neural network classifiers. International journal of computer applications. 2017;175(4):7-9.

[11] scikit-learn developers. 1.13. Feature selection; 2024. Available from: `https://scikit-learn.org/stable/modules/feature_selection.html`.

[12] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003;3(Mar):1157-82.

[13] Vincent AM, Jidesh P. An improved hyperparameter optimization framework for AutoML systems using evolutionary algorithms. Scientific Reports. 2023 3;13:4737.

[14] Bischl B, Binder M, Lang M, Pielok T, Richter J, Coors S, et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. WIREs Data Mining and Knowledge Discovery. 2023 3;13.

[15] Pilario KE, Cao Y, Shafiee M. A Kernel Design Approach to Improve Kernel Subspace Identification. IEEE Transactions on Industrial Electronics. 2020 05;PP:1-1.

[16] Bradshaw TJ, Huemann Z, Hu J, Rahmim A. A Guide to Cross-Validation for Artificial Intelligence in Medical Imaging. Radiology: Artificial Intelligence. 2023;5(4):e220232.

[17] Subramanian J, Simon R. Overfitting in prediction models – Is it a problem only in high dimensions? Contemporary Clinical Trials. 2013;36(2):636-41.

[18] Erickson B, Kitamura F. Magician's Corner: 9. Performance Metrics for Machine LearningModels. Radiology: Artificial Intelligence. 2021;3(3):e200126.

[19] Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. Journal of clinical epidemiology. 2003;56(5):441-7.

[20] Hosseini M, Powell M, Collins J, Callahan-Flintoft C, Jones W, Bowman H, et al. I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data. Neuroscience & Biobehavioral Reviews. 2020;119:456-67.

Part III

Clinical applications

Application of the technologies described in part II on clinical data.

Chapter 10

Statistical parametric mapping for segmentation evaluation

A comparison experiment for the segmentation accuracy of knee MRI is described in this chapter. Four state-of-the-art neural network architectures are trained with identical hyperparameters (such as patch size, batch size, augmentation techniques, optimizers,..). Next, these four trained neural networks were tested on a separate test set of MRI scans, that had not been previously seen by the model. Standard segmentation accuracy performance metrics are compared, in addition to a statistical parametric mapping of the eventual segmentation errors on the mean 3D surface of the anatomy of interest. This allows for the visualization of the largest segmentation errors by anatomical location and helps reveal overall trends in the over- or underestimation of certain anatomical structure volumes. Ultimately, leveraging expert knowledge relevant to the final anticipated application of the 3D models, the most appropriate model for the task can be selected.

This chapter is adopted from a manuscript submitted at NEJM AI:
Grammens, J., Danckaers, F., Van Haver, A., Verdonk, P., Sijbers, J. From Dice-scores to anatomical relevance: a comparative study of four state-of-the-art neural networks for automated knee MRI segmentation.

10.1 Introduction

Personalized medicine is reshaping the way pathologies are diagnosed and treated. It aims to provide tailored care based on an individual's unique genotype and phenotype [1, 2]. The integration of advanced 3D imaging into standard clinical practice, such as magnetic resonance imaging (MRI) and computed tomography (CT) provides detailed images of the human body and the patient's disease state. Not only the visual aspect of a lesion (signal intensity), but also the three-dimensional size of it and its relation to the surrounding anatomical structures can be assessed [3]. By reconstructing 3D models from scan data, healthcare professionals can visualize complex relationships between organs, tissues, and pathologies. This is particularly valuable in surgical planning, where surgeons can virtually simulate procedures and identify potential challenges before entering the operating room.

These 3D reconstructions only recently began to find entrance into the routine clinical care. Thanks to advancements in computer vision and artificial intelligence, the process of anatomy and/or lesion detection in scan data could be automated [4]. Prior to automation, the manual segmentation of scan data was a laborious and massively time-consuming process that significantly impacted operational efficiency and cost-effectiveness [5]. This task involved manually annotating specific regions of interest within the scan data, often requiring meticulous attention to detail and expert knowledge. The repetitive nature of this process, coupled with the potential for human error, made it a substantial bottleneck in the workflow.

Computer vision algorithms, accelerated by the parallel processing capabilities of Graphics Processing Units (GPUs), have facilitated substantial advancements in automated segmentation techniques. The next step before massive clinical integration is a thorough validation of these innovations. While the commonly employed metrics (such as the Dice or Jacquard index) provide valuable insights, they may not fully capture the nuances of medical applications [6]. Moreover, their use can often lead to misleading or obscure results when misapplied or misinterpreted, undermining their intended purpose.

One issue arises when accuracy metrics are averaged over multiple anatomical structures. While averaging can provide a general sense of an algorithm's performance, it often masks significant variations across different cases. For instance, an algorithm that performs well on common, easily segmented tissues might perform poorly on rare or complex structures. When these results are averaged, the overall accuracy may appear reasonable, even though the algorithm fails for critical regions of interest. This can lead to overly optimistic assessments that do not reflect the true utility of the segmentation method for the application at hand. The use of distinct and often incomparable metrics further complicates the evaluation landscape [7]. Even when the same metric is used, differences in calculation methods (such as micro versus macro aggregation) can lead to different interpretations. Micro-averaging gives equal weight to each pixel or voxel, which can skew results if the classes are imbalanced. In contrast, macro-averaging treats each class or anatomy equally, which might underemphasize performance on larger objects. When comparing results across studies or datasets, these differences in metric calculation can lead to confusion and misinterpretation. Furthermore, the relationship between the size of the region of interest (ROI) and the overall image size can distort performance metrics. In cases where the ROI is small relative to the entire image, a high accuracy score might simply reflect the algorithm's ability to correctly label the majority background class, while failing to capture the ROI with sufficient detail. This issue is particularly problematic when the structures of interest, such as small lesions or anatomical boundaries, occupy only a tiny fraction of the image. Additionally, the distribution of latent variables in the training, validation, and test sets plays a crucial role in the generalizability of segmentation models [8]. If these sets do not share a similar distribution of these variables—such as patient demographics, scanner types, or anatomical variations—the performance metrics may not accurately reflect how the algorithm will perform in real-world settings. This is particularly problematic when the test set is not representative of the diversity encountered in clinical practice, leading to an overestimation of the algorithm's robustness. A final critical shortcoming of traditional segmentation metrics is their lack of spatial context. Metrics such as Dice coefficient, Jaccard index, Average Surface Deviation or Hausdorff distance treat all segmentation errors equally [7], regardless of

their anatomical or spatial location. However, errors in critical areas (e.g., the boundary of a tumor) may be far more consequential than errors in less significant regions. The absence of spatial weighting in these metrics means that potentially critical errors can go unnoticed, leading to an overestimation of the segmentation algorithm's clinical applicability.

In this chapter, we implemented a method to bridge this significant gap by including the spatial context during evaluation of segmentation accuracy. More precisely, segmentation errors will be assessed in the 3D model domain, where the final clinical application of the segmentation usually lies. A statistical parametric mapping enables the visualization of point-wise (or anatomical landmark-wise) descriptive statistics for eventual 3D surface deviations from the ground truth. This way, the clinical importance of segmentation inaccuracies can be assessed. For example, a systematic over- or underestimation of knee cartilage thickness might be more critical in the clinical decision process than some inaccuracies in the non weight-bearing regions.

10.2 Materials and methods

All model training and validation experiments were performed in customly adapted Python scripts, within the nnUNet framework (v2) [9]. Thereby, following python packages were used during the complete workflow: pytorch [10], MONAI [11], VTK [12], pyvista[13] and nnUNet [9]. All model training and inference was performed on a desktop workstation (AMD® Ryzen™ 9 7900x, 128 GB RAM and a NVIDIA® GeForce® RTX 4090 GPU).

10.2.1 Dataset description

The dataset for model training and evaluation was the Osteoarthritis Initiative Database [14]: a large open-source database containing longitudinal follow-up data related to knee osteoarthritis (clinical and radiological) from 4970 individuals. Three large cohorts were present: the progression, the incidence and the control group. The progression cohort was already diagnosed with knee OA (both radiographic signs and frequent symptoms, n=1389) at baseline, while the incidence group was at risk of developing knee OA (frequent knee symptoms without radiographic signs, risk factors present, n=3285) and the control cohort had neither radiographic nor symptomatic knee OA (n=122) over the complete 8 years of follow-up. 507 manual segmentations containing annotated femoral bone, femoral cartilage, tibial bone and tibial cartilage served as ground truth for training and model performance evaluation [15].

10.2.2 Image preprocessing

This dataset was split into a training (N=253) and a test set (N=254), respecting an equal distribution of distinct OA grades (Kellgren-Lawrence [16]) in both splits. All image preprocessing was handled by the nnunetv2 framework [9]. Only one configuration was trained and tested, which resampled the images into an isotropic resolution

(0.7mm $\times$ 0.7mm $\times$ 0.7mm) matching the largest voxel dimension of the original scan volume (mediolaterally). Downsampling the finest voxel dimensions was preferred to upsampling the most coarse dimension, due to GPU memory restrictions. Additionally, the patch size of $128 \times 128 \times 128$ was chosen in order to fit the following requirements: equal three-dimensional size and a batch size of 2 feasible for training on a GPU with 24GB VRAM. Patches were eventually cut from a resampled $160 \times 199 \times 199$ scan volume, with an overlap of 75% or tile step size of 32.

10.2.3 Model training

Four state-of-the-art network architectures, each with their advantages and challenges, were selected to apply on the task of knee MRI segmentation: nnU-Net (with residual encoding) [9], MedNeXt [17], Swin-UNetR [18] and SegMamba [19]. The first two are purely convolutional networks, while the last two are hybrid approaches combining some elements from convolutional and Transformer networks. Model training for all four networks was performed in the nnU-Net framework, to ensure identical or at least comparable (hyper)parameters during training. All models were trained in a 5-fold cross-validation experiment on the training set, ensuring identical data splits for all four network setups. For all the networks that were assessed, training was performed for 1000 epochs or until the learning curve showed clear signs of overfitting (trend to further decreasing training loss in the context of a status quo or even increasing validation loss). Stochastic gradient descent (momentum=0.99) was used as optimizer. The learning rate was scheduled by a cosine annealing scheme with warm restarts as shown in Figure 7.1. The used loss function during training was the combo loss with an equal contribution of the Dice and cross entropy loss function. A preliminary best model checkpoint was saved when the exponential moving average of Dice-scores (previous value $\times 0.9$ + new score $\times 0.1$) for the foreground classes in the validation set exceeded the previous best exponential moving average. Use of GPU memory by the training script was collected from the GPU driver System Management Interface.

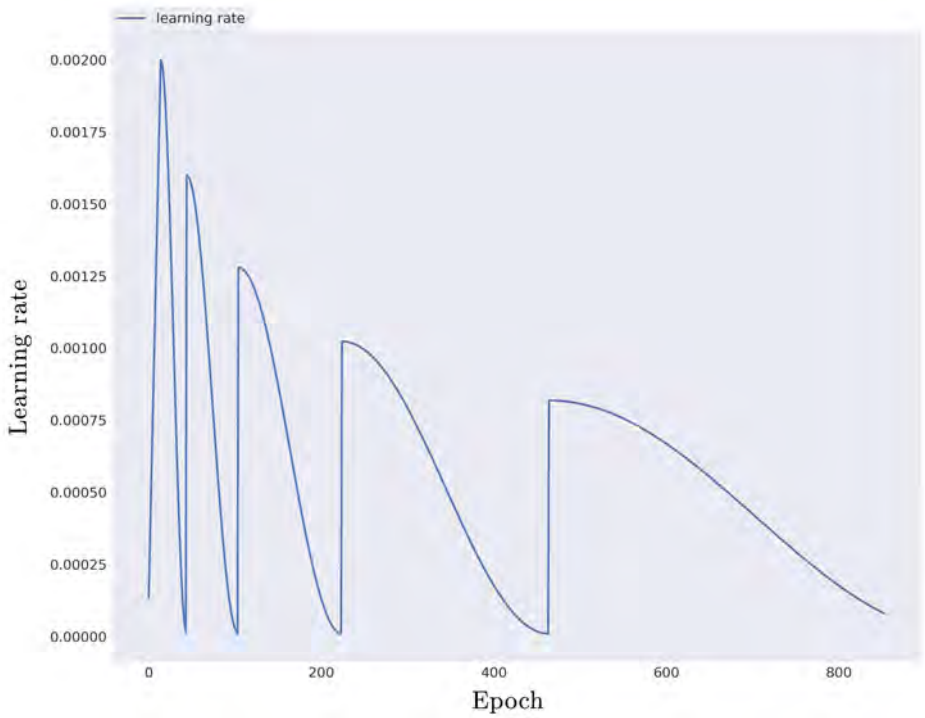


Figure 10.1: Learning rate as scheduled during the experiments. A linear warmup was applied, followed by a cosine annealing schedule with warm restarts. The learning rate for every next warm restart was reduced to 80% of the previous warm restart.

10.2.4 Model inference

Gaussian weighted aggregation of overlapping patches (75%) was employed for validation and inference purposes. An ensemble of the 5-fold cross-validation models was finally used to perform inference on all MRI scans in the test set (N=254).

10.2.5 Conventional segmentation accuracy metrics

Following metrics were used to assess differences between the model predictions and the manual ground truth segmentations: Dice similarity coefficient (DSC), average surface deviation (ASD, calculated in the image domain) and Hausdorff distance [7]. Inference time and GPU memory usage, as reported by the GPU driver's System Management Interface, were recorded to give some insights into practical feasibility on other hardware configurations.

10.2.6 3D post-processing

As the anatomy of interest only consist of coherent regions, small isolated particles (as a result of noise) were removed, depending on the assessment per anatomy of interest by the nnUNet-framework on the training dataset. The obtained segmentation masks from both ground truth and network inference were processed through a marching cubes algorithm to create 3D surface models in the following configurations: femoral bone, femoral bone with femoral cartilage, tibial bone and tibial bone with tibial cartilage. No separate cartilage 3D models were created, as this would pose an additional challenge to the surface registration step. Osteoarthritis involves the degeneration and wear of cartilage and could therefore require a change in topology in the 3D meshes, depending on the amount of full-thickness cartilage defects (or simpler described: holes in the shell-like 3D cartilage models). In that case, surface registration is nearly impossible for that region, as no corresponding anatomical landmarks could be placed where the cartilage is completely gone. The resulting 3D surface models showed some staircase effects, originating from the original voxel geometry. A Taubin volume-preserving smoothing step was therefore performed to create smooth, more realistic 3D surfaces. Finally, to ensure a homogenous mesh resolution and level of detail over the complete 3D surface, an isotropic remeshing step was performed to ensure all 3D models consist of 40,000 vertices.

10.2.7 Surface registration

The 3D models originating from the ground truth segmentations were in a next step registered [20]. That is, for each point on the 3D surface of a template case (typically randomly chosen or mean shape) the anatomically corresponding point on each 3D surface is searched for (see ”section 11.3. Surface registration” for a detailed description of this process). The goal is to obtain a dense set of anatomically corresponding pseudo-landmarks from all cases in the dataset, so that the segmentation errors can be aggregated per anatomical location of interest. The segmentation errors are assessed by distance maps between the 3D surface models of the registered ground truth models and the post-processed output of the neural networks. These distance maps are then aggregated point-wise (in anatomical correspondence) over the test set cases to visualize the descriptive statistics, including the maximal absolute surface deviation (robustness) and mean absolute value surface deviation (general trends in accuracy).

10.2.8 Descriptive statistics of segmentation errors

Mean and maximal absolute values of surface deviations from ground truth segmentations were computed across the entire test set to assess eventual general segmentation errors and method robustness. The mean absolute error was chosen for its simplicity and robustness against outliers [21, 22] in comparison with the also commonly used root mean square error (RMSE).

10.3 Results

10.3.1 Conventional segmentation accuracy metrics

The Dice Similarity Coefficient (DSC), used to evaluate spatial overlap, indicated that the nnUNet achieved the highest performance for the femoral bone, femoral cartilage, and tibial bone. For the tibial cartilage, however, SegMamba demonstrated superior DSC values. Metrics based on spatial distance, specifically the Average Surface Distance (ASD) and Hausdorff Distance (HD), consistently identified nnUNet as the best-performing method across all anatomical structures examined (femoral bone, femoral cartilage, tibial bone, and tibial cartilage). A comprehensive summary of the conventional numeric accuracy metrics from this experiment is provided in Table 7.1.

Table 10.1: Conventional segmentation metric results from this experiment: mean +/- standard deviation (SD).

Network architecture	DSC				ASD [mm]				HD [mm]			
	FB	FC	TB	TC	FB	FC	TB	TC	FB	FC	TB	TC
nnUNet	**0.986**	**0.905**	**0.988**	0.856	**0.229**	**0.221**	**0.201**	**0.245**	**2.74**	**4.85**	**2.54**	**4.47**
MedNeXt	0.986	0.898	0.987	0.855	0.239	0.241	0.214	0.251	2.94	4.96	2.73	4.56
SwinUNetR	0.985	0.898	0.987	0.859	0.244	0.240	0.210	0.245	3.15	5.10	2.74	4.65
SegMamba	0.986	0.901	0.987	**0.860**	0.242	0.233	0.207	0.246	3.21	5.03	2.70	4.64

DSC: Dice similarity coefficient, ASD: average surface deviation (calculated in voxel domain), HD: Hausdorff distance, FB: femoral bone, FC: femoral cartilage, TB: tibial bone, TC: tibial cartilage

10.3.2 3D surface mapping

Femoral bone

All network architectures exhibited the highest mean absolute error in the anterior region above the femoral trochlea (highlighted in red in Figure 7.2). Additionally, the posteromedial region of the femoral notch also showed a higher tendency for surface deviations from the ground truth.

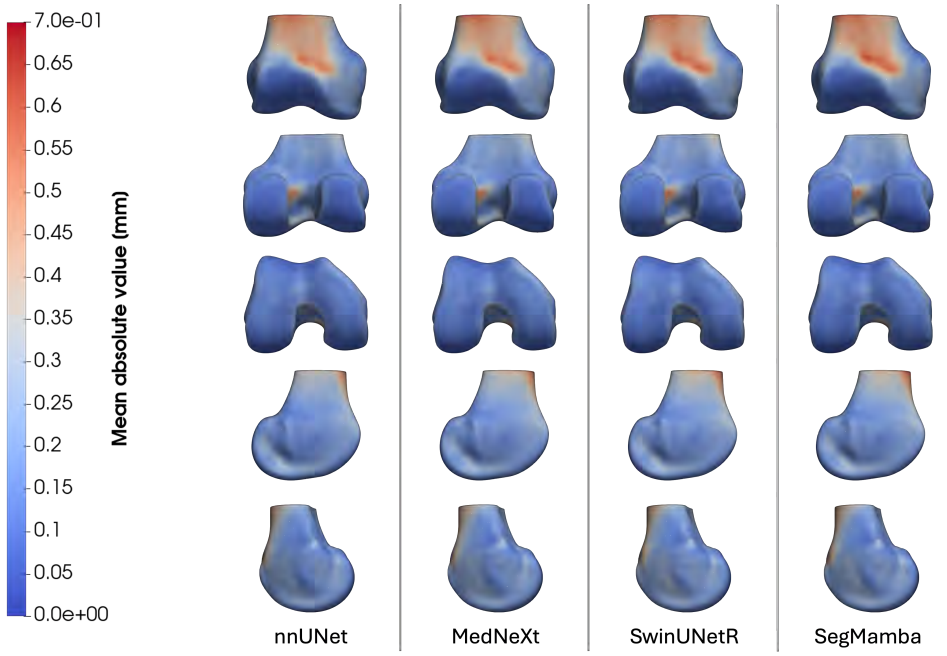


Figure 10.2: Distance map between manual and neural network segmentations of 3D femoral bone surfaces, visualized as the pointwise mean absolute error plotted on the mean 3D mesh of femoral bone. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the femoral bone.

Figure 7.3 highlights the maximal absolute value errors. Unlike nnUNet and MedNeXt, SwinUNetR and SegMamba exhibit additional red regions at the anteromedial trochlear border, a common osteophyte location in patellofemoral osteoarthritis. Moreover, notable maximal segmentation errors were observed posterior to the lateral femoral epicondyle and around the medial epicondyle. The distal lateral condyle showed varying levels of maximal absolute segmentation error across methods, with SegMamba exhibiting the highest error across all test cases.

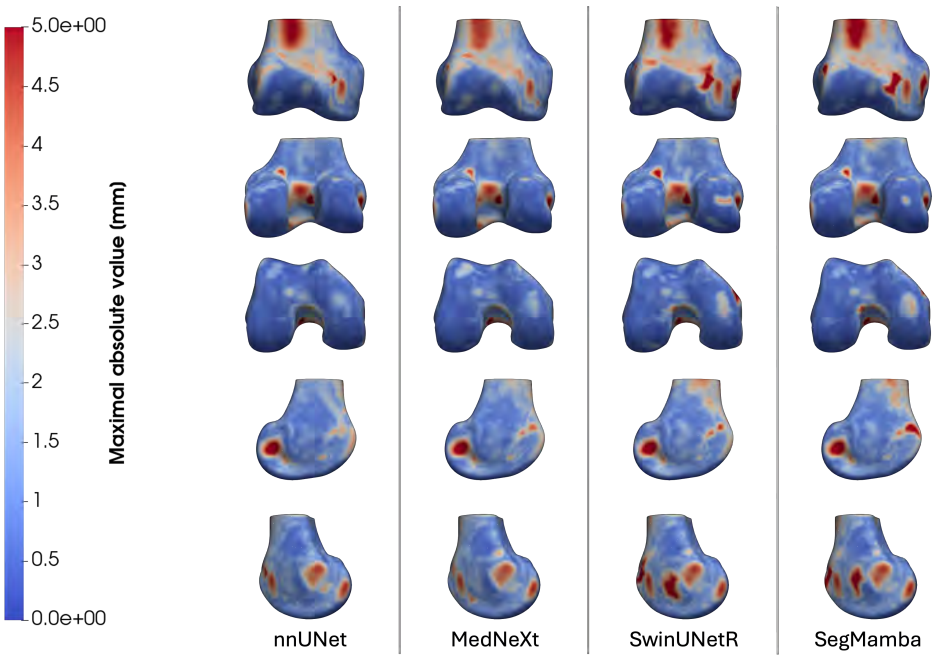


Figure 10.3: Distance map between manual and neural network segmentations of 3D femoral bone surfaces, visualized as the pointwise maximal absolute error plotted on the mean 3D mesh of femoral bone. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the femoral bone.

Femoral bone and cartilage

Unlike the conventional segmentation accuracy metrics described in subsection 7.3.1: Conventional segmentation accuracy metrics, 3D segmentation error mapping was performed on the combined 3D model femoral bone and cartilage, and not just the cartilage. As bone surface segmentation errors were discussed previously, this analysis focuses on the cartilage surface. The mean absolute segmentation error revealed that cartilage border regions, especially in the anterior (patellofemoral) compartment, were more susceptible to segmentation errors. nnUNet exhibited slightly better performance at the anterior part of the distal facet in the lateral condyle, the posterior medial condyle, and anteromedial border of the medial condyle (Figure 7.4).

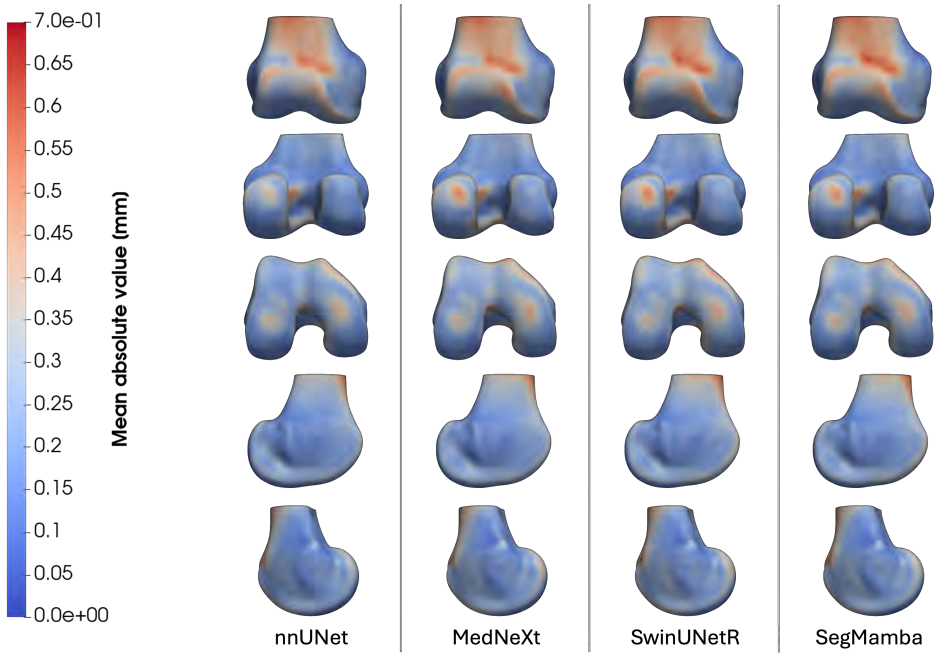


Figure 10.4: Distance map between manual and neural network segmentations of 3D femoral bone and cartilage surfaces, visualized as the pointwise mean absolute error plotted on the mean 3D mesh of femoral bone and cartilage. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the femoral bone and cartilage.

The maximal absolute segmentation error revealed no significant large errors within the cartilage region. Only a minor red spot, representing a larger maximal segmentation error, was observed at the posterior lateral condyle in the SwinUNetR results (Figure 7.5).

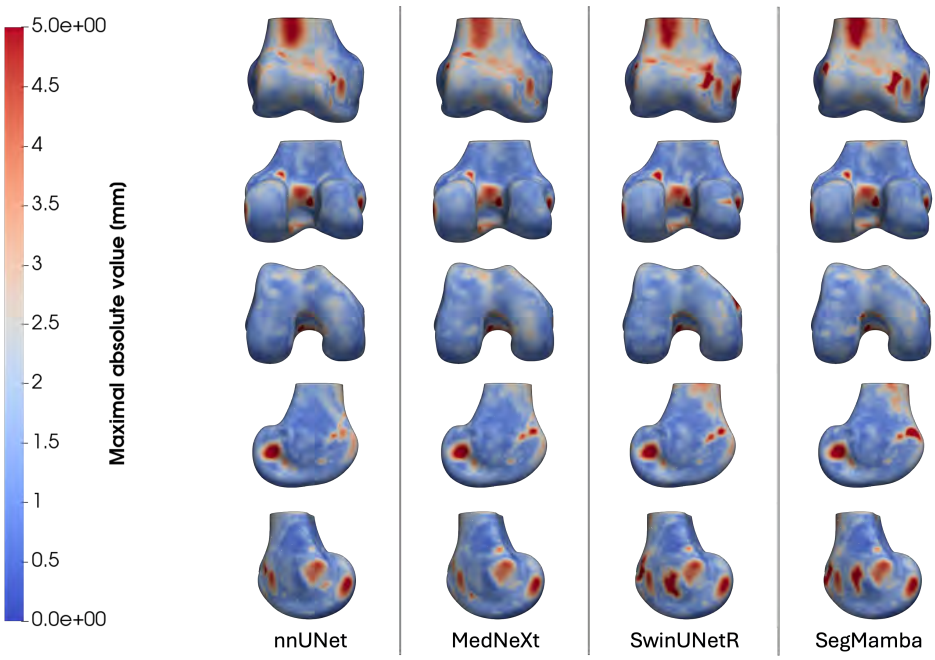


Figure 10.5: Distance map between manual and neural network segmentations of 3D femoral bone and cartilage surfaces, visualized as the pointwise maximal absolute error plotted on the mean 3D mesh of femoral bone and cartilage. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the femoral bone and cartilage.

Tibial bone

For the tibial bone, the tuberositas tibiae exhibited the highest mean absolute segmentation error. Slightly elevated errors were observed at the tibial spines and the anterior region. No remarkable differences were found between the four methods for these anatomical regions (Figure 7.6).

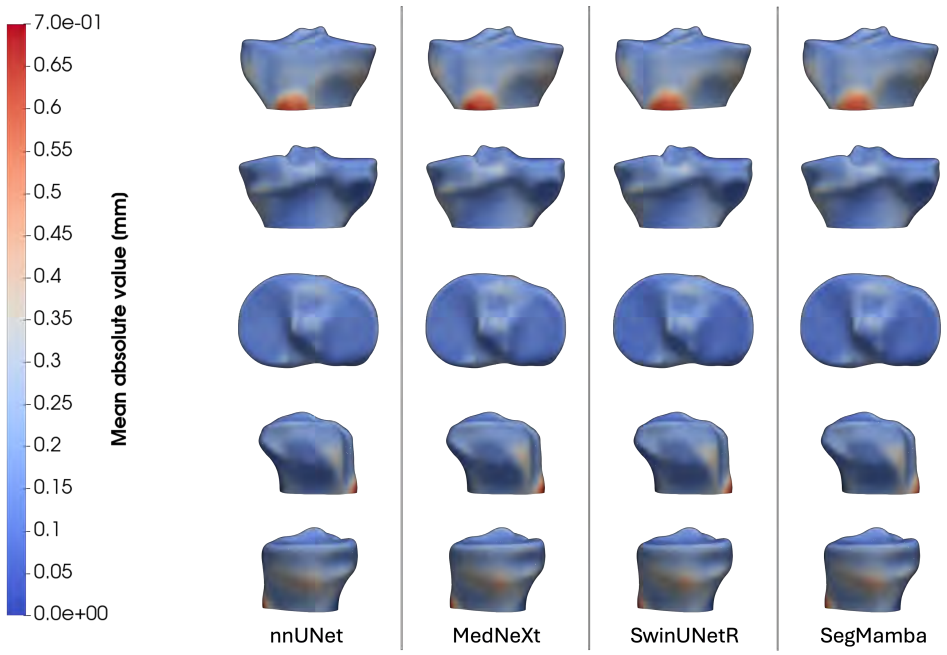


Figure 10.6: Distance map between manual and neural network segmentations of 3D tibial bone surfaces, visualized as the pointwise mean absolute error plotted on the mean 3D mesh of tibial bone. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the tibial bone.

The maximal absolute value error was again high at the region of the tuberositas tibiae for all four network architectures (Figure 7.7. In addition, a red spot is observed anterior of the tibial spines, at the approximate insertion region of the anterior cruciate ligaments, this is the least pronounced for the output of the MedNeXt network. The SegMamba network is the only method with a remarkable maximal absolute value error at the medial side of the tibial plateau. A red spot is also observed at the posterolateral ridge of the tibial plateau for all networks, in different grades: it was maximal in the nnUNet and the MedNeXt and least in the SwinUNetR network output. All four network architectures struggled with accuracy for at least one case at the tibial tuberosity, demonstrating consistently high maximal absolute value errors. A common error pattern emerged, with a red spot appearing anterior to the tibial spines, likely corresponding to the anterior cruciate ligament insertion. This error was most pronounced for nnUNet and MedNeXt, while MedNeXt showed the least remarkable error in this region. The SegMamba network was the only model to exhibit a notable maximal absolute value error at the medial side of the tibial plateau. Additionally, a red spot was observed at the posterolateral tibial plateau ridge for all networks. While this error was present across all models, its severity varied: nnUNet and MedNeXt displayed the most prominent errors in this region (error magnitude and region size), while SwinUNetR demonstrated the least.

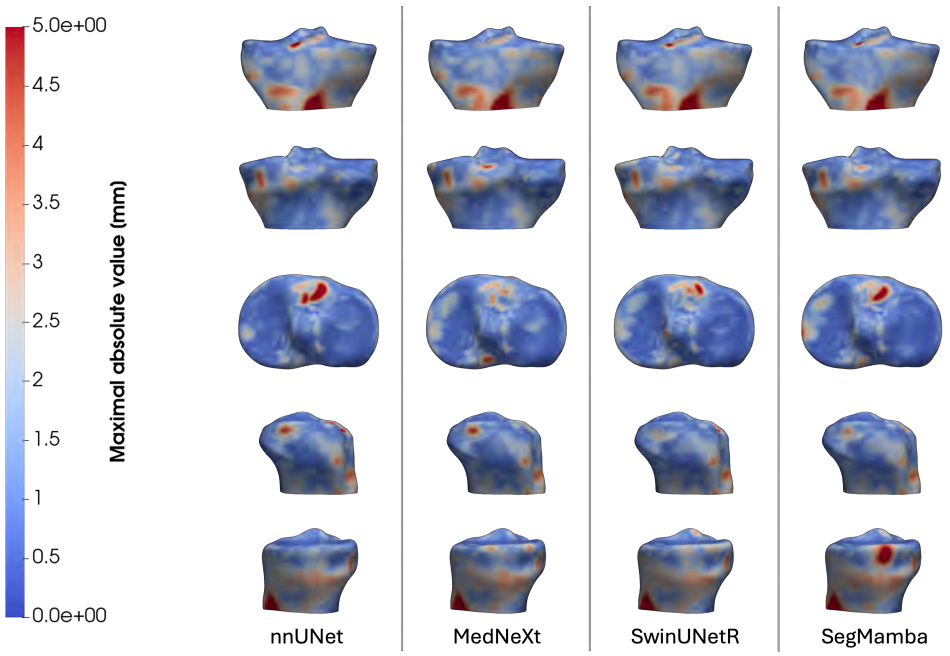


Figure 10.7: Distance map between manual and neural network segmentations of 3D tibial bone surfaces, visualized as the pointwise maximal absolute error plotted on the mean 3D mesh of tibial bone. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the tibial bone.

Tibial bone and cartilage

As bone surface results were discussed previously, this analysis focuses solely on the cartilage surface, visualized by the top view (middle row in Figure 7.8). Absolute value errors were generally higher at cartilage borders and central tibiofemoral contact points. Additionally, the tibial spines exhibited a tendency towards higher absolute value segmentation errors.

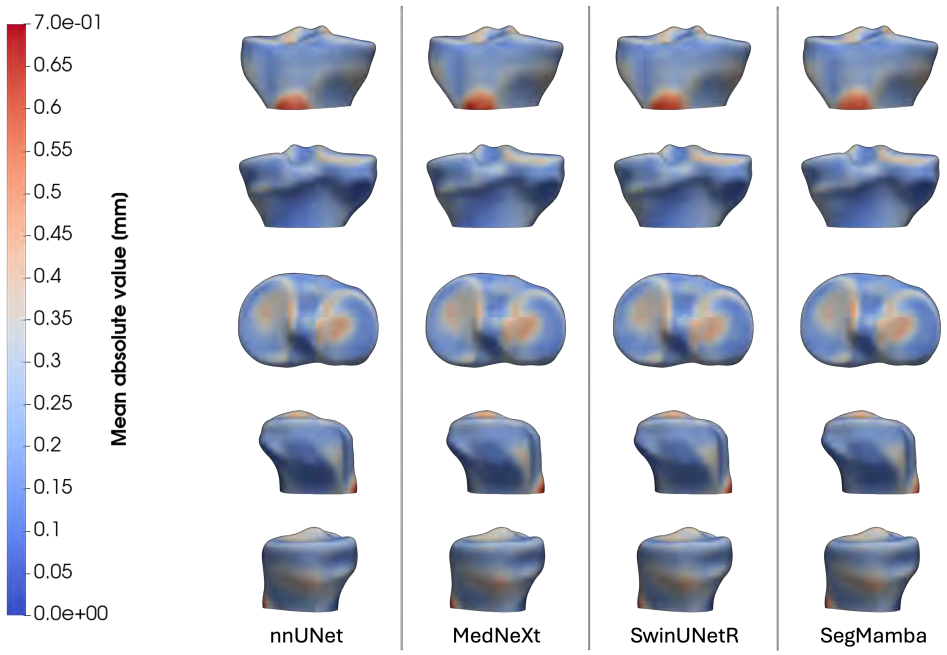


Figure 10.8: Distance map between manual and neural network segmentations of 3D tibial bone and cartilage surfaces, visualized as the pointwise mean absolute error plotted on the mean 3D mesh of tibial bone and cartilage. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the tibial bone and cartilage.

The maximal absolute segmentation errors, where slightly higher for the cartilage layer (lighter blue) than for the tibial bone at the tibial plateau (7.3.2). Maximal absolute value segmentation errors were in the same range for all four networks, with a tendency to be smallest for the output of the nnUNet network (Figure 7.9).

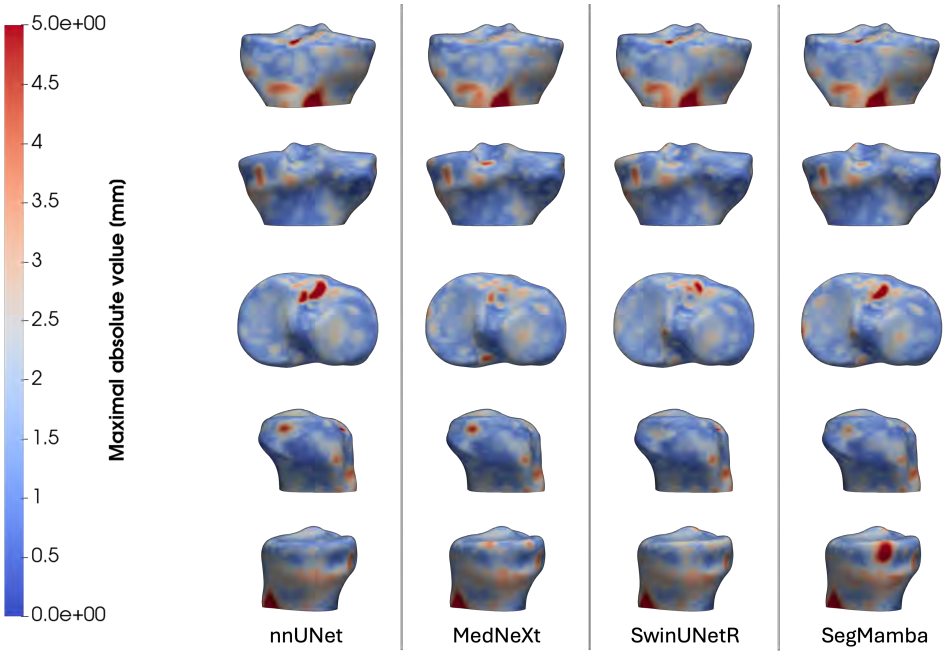


Figure 10.9: Distance map between manual and neural network segmentations of 3D tibial bone and cartilage surfaces, visualized as the pointwise maximal absolute error plotted on the mean 3D mesh of tibial bone and cartilage. The columns from left to right contain the results from the nnUNet, MedNeXt, SwinUNetR and Segmamba neural networks. The rows from top to bottom contain anterior, posterior, distal, lateral and medial views of the tibial bone and cartilage.

10.3.3 Hardware requirements and inference time

The amount of GPU memory used during training and inference is summarized in Table 7.2, as well as the time needed to run inference on 1 resampled scan volume. For the same patch and batch sizes, the nnUNet architecture requires the least amount of hardware resources and computational time for inference.

Table 10.2: Conventional segmentation metrics for image segmentation.

Network architecture	GPU memory usage during training [MiB]	GPU memory usage during inference [MiB]	Inference time [s]
nnUNet	9 304	2 774	3.91
MedNeXt	16 378	4 898	9.10
SwinUNetR	19 198	5 978	8.58
SegMamba	23 026	3 720	10.73

GPU: graphics processing unit, MiB: mebibyte (equals $2^20 = 1048576$ bytes), s: seconds

10.4 Discussion

In this study, we evaluated the performance of four neural network architectures—nnUNet, MedNeXt, SwinUNetR, and SegMamba—on the task of 3D MRI knee scan segmentation. Our findings indicate only subtle differences across these models, both in terms of conventional accuracy metrics as in terms of 3D mapped mean and maximal absolute errors. This uniformity in performance indicates that, for applications where this subtle variation in segmentation accuracy is acceptable, other factors such as ease of use, customization, and easy to fulfill hardware requirements should be prioritized when choosing a model. Our results notably favor nnUNet (with residual encodings) in this regard, primarily due to its user-friendly configuration pipeline, fast inference time and low GPU memory use during training and inference. The basic U-Net architecture integrates seamlessly within the nnUNet framework, and this combination has been proven advantageous due to its automated configuration system, which reduces the burden of manual hyperparameter tuning, a common challenge in deep learning implementations for medical imaging [9, 23].

Of note, the achieved segmentation accuracy was obtained without extensive post-processing techniques. In the original experiment on the same dataset by Ambellan et al. [15], similar segmentation accuracies where achieved by less optimized convolutional neural networks, although through a multi-step approach including a final postprocessing step involving regularization by a statistical shape model. They obtained a DSC score of 0.986 for the FB, 0.899 for the FC, 0.986 for the TB and 0.856 for the TC. A more recent study on the same dataset, employing a Swin-UNetR for segmentation of femoral and tibial bone, reported a DSC score of 0.986 for the FB and 0.987 for the TB [24], which is comparable to our results. Our results also align with the study of Isensee et al [8], comparing multiple state-of-the-art model architectures for multiple segmentation tasks. None of the proposed novel model architectures resulted in a remarkably improvement upon the nn-UNet baseline. They suggested a potential validation bias in the articles introducing these new neural network architectures for computer vision. A reliable and generalizable validation experiment requires several factors, including an optimally configured (hyperparameters) baseline model, a sufficiently large and diverse test dataset, and comparable train/validate/test splits (both in terms of quantity and quality) in the comparative experiments. Unfortunately, some of these crucial elements are sometimes overlooked, potentially resulting in overly optimistic validation results [8].

Beyond demonstrating that none of the chosen neural network architectures are clearly superior, this study also sheds light on anatomical regions most susceptible to segmentation errors. Although the literature on manual knee MRI segmentation accuracy is scarce [25, 26], the regions with higher observed segmentation errors align with our experience (Figures 7.10, 7.11 and 7.12). Manual segmentation is prone to intra- and interobserver errors, limiting its reproducibility to a certain extent. Despite this limitation, manual segmentations are still considered ground truth. Our study suggests that, within the constraints of a 24 GB GPU memory budget, it may be challenging to significantly improve upon the achieved segmentation accuracies. As accuracy gains by novel network architecture innovations become more and more difficult to achieve, an attention shift towards quality control of images and ground truth segmentations [27] is desirable. Intuitively, an upper limit for network

performance is posed by both the size (and included variation) and quality of the training dataset.

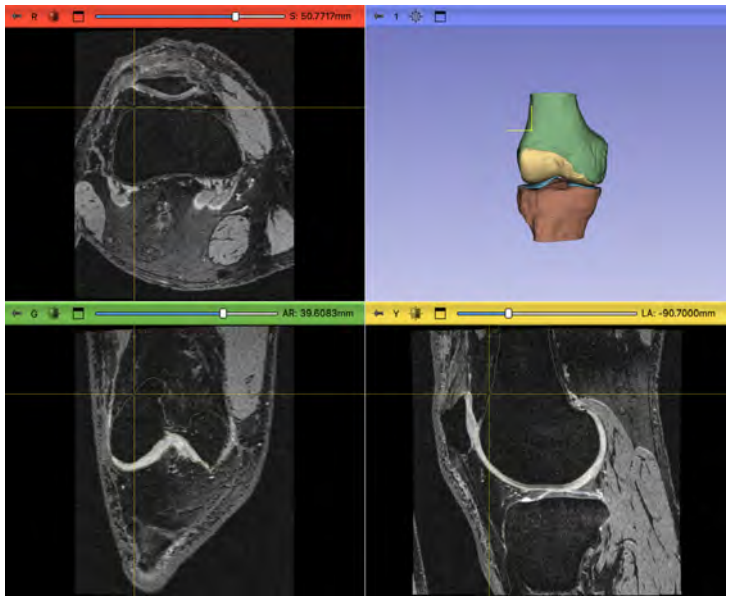


Figure 10.10: Screenshot from the 3D Slicer software package with the crosshair at the anterior femoral bone shaft. The low tissue contrast between cortical bone and surrounding soft tissue makes the manual delineation of the anterodistal aspect of the femoral shaft challenging.

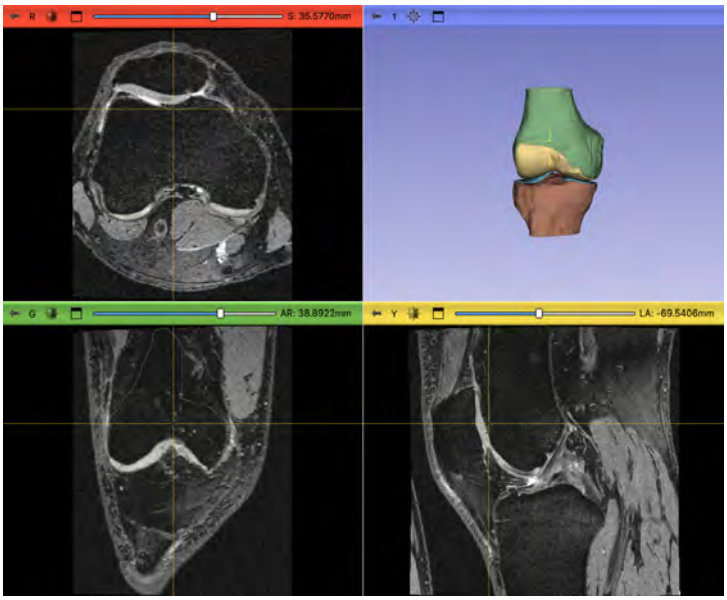


Figure 10.11: Screenshot from the 3D Slicer software package with the crosshair at the anteromedial border of the trochlea. The low tissue contrast between cortical bone and surrounding soft tissue renders the manual delineation of the anteromedial border of the femoral trochlea challenging.

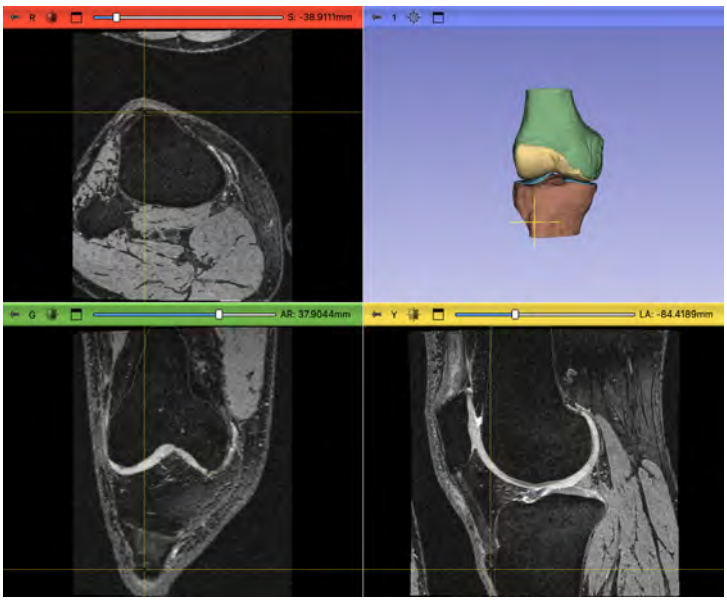


Figure 10.12: Screenshot from the 3D Slicer software package with the crosshair at the tuberositas tibiae. Remark the low tissue contrast between cortical bone and the surrounding soft tissue, leading to uncertainty to delineate manually the tibial cortical bone of the tuberositas tibiae. Both the insertion of the quadriceps tendon at the tuberositas tibiae and the cortical bone have a low MRI signal in these 3D DESS images.

While our training and test datasets were balanced with respect to the Kellgren-Lawrence OA grades in train and test set, a potential limitation of our study is the lack of quantitative osteophyte size and location description, which could introduce potential latent distribution discrepancies between training and validation sets. Additionally, to achieve isotropic input scan resolution while minimizing GPU memory usage, images were downsampled during preprocessing, which could potentially reduce resolution through aliasing effects. However, as this downsampling was limited to the sagittal plane, reducing resolution from approximately 0.35×0.35 mm to 0.7×0.7 mm, it is unlikely that this change meaningfully impacted 3D model detail in the context of our specific task. The smallest anatomical structures analyzed, such as femoral and tibial cartilage layers, naturally exhibit time- and load-dependent thickness variations [28], independent of measurement technique. Thus, a 3D model accuracy of 0.7 mm was deemed sufficient for most future clinical applications. It is also noteworthy that these findings were derived from specific 3D DESS MRI scans, yielding near-isotropic, high signal to noise ratio (SNR) images with a clear contrast between the intra-articular synovial fluid and the cartilage [29]. However, these 3D MRI protocols are not yet widely implemented in routine clinical knee imaging, but only begin to enter clinical practice [30]. Finally, regarding GPU memory usage, the variation in GPU requirements across the four network architectures was substantial. For example, the least resource-intensive network, nnUNet, did not fully utilize the available GPU memory. To isolate the effects of network architecture, all other parameters were held constant. Consequently, neither larger patch sizes nor batch sizes were used during training, potentially limiting maximal model performance.

The effectiveness of a segmentation algorithm is fundamentally influenced by two key components: the algorithm itself and the dataset of images with corresponding ground truth segmentations. While extensive research has focused on developing and refining algorithms, it is crucial not to overlook the quality and volume of the data used. Recent studies have demonstrated the potential to predict segmentation accuracy in real-time, even in the absence of ground truth data, offering a possible method for evaluating image quality and determining whether the proposed algorithm can effectively segment the image [31, 32]. Interestingly, the emphasis on large training datasets may be overstated [33], while greater attention to improving the accuracy of ground truth segmentations could be an interesting future research direction [34, 35, 36].

10.5 Conclusion

In conclusion, our study underscores that the specific choice of neural network architecture from a set of state-of-the-art networks (nnUNet, MedNeXt, SwinUNetR, and SegMamba) has limited impact on segmentation accuracy in MRI knee scans . Instead, practical considerations such as deployment and ease of use, favoring architectures like nnUNet, should guide model selection in clinical and translational research settings. Additionally, our findings align with other studies, suggesting a critical role for the training dataset quality. Future improvements in segmentation accuracy may therefore be more readily achieved by focusing on both image and label quality than by experimenting with increasingly complex architectures. At last, the adage 'garbage in, garbage out' continues to hold true in the field of machine learning...

Bibliography

[1] Gameiro GR, Sinkunas V, Liguori GR, Auler-Júnior JOC. Precision Medicine: Changing the way we think about healthcare. Clinics. 2018;73:e723.

[2] Wang Z, Huang J, Xie D, He D, Lu A, Liang C. Toward Overcoming Treatment Failure in Rheumatoid Arthritis. Frontiers in Immunology. 2021 12;12.

[3] Sadeghi MH, Sina S, Alavi M, Giammarile F, Yeong CH. PET/CT-based 3D multi-class semantic segmentation of ovarian cancer and the stability of the extracted radiomics features. Physical and engineering sciences in medicine. 2024 9.

[4] Berghe TVD, Babin D, Chen M, Callens M, Brack D, Maes H, et al. Neural network algorithm for detection of erosions and ankylosis on CT of the sacroiliac joints: multicentre development and validation of diagnostic accuracy. European radiology. 2023 11;33:8310-23.

[5] Ungureanu G, Serban LN, Beni L, Florian SI. Enhancing Patient Comprehension in Skull-Base Meningioma Surgery through 3D Volumetric Reconstructions: A Cost-Effective Approach. Journal of personalized medicine. 2024 9;14.

[6] Hirling D, Tasnadi E, Caicedo J, Caroprese MV, Sjögren R, Aubreville M, et al. Segmentation metric misinterpretations in bioimage analysis. Nature Methods. 2024 2;21:213-6.

[7] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Medical Imaging. 2015 12;15:29.

[8] Isensee F, Wald T, Ulrich C, Baumgartner M, Roy S, Maier-Hein K, et al.. nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation; 2024. Available from: `https://arxiv.org/abs/2404.09556`.

[9] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2021;18(2):203-11.

[10] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al.. PyTorch: An Imperative Style, High-Performance Deep Learning Library; 2019. Available from: `https://arxiv.org/abs/1912.01703`.

[11] Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al.. MONAI: An open-source framework for deep learning in healthcare; 2022. Available from: `https://arxiv.org/abs/2211.02701`.

[12] Schroeder W, Martin K, Lorensen B. The Visualization Toolkit (4th ed.). Kitware; 2006.

[13] Sullivan CB, Kaszynski A. PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). Journal of Open Source Software. 2019 may;4(37):1450.

[14] Lester G. Clinical research in OA - The NIH Osteoarthritis Initiative. Journal of musculoskeletal & neuronal interactions. 2008;8:313-4.

[15] Ambellan F, Tack A, Ehlke M, Zachow S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. Medical Image Analysis. 2019.

[16] Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. Annals of the Rheumatic Diseases. 1957;16(4):494-502.

[17] Roy S, Koehler G, Ulrich C, Baumgartner M, Petersen J, Isensee F, et al.. MedNeXt: Transformer-driven Scaling of ConvNets for Medical Image Segmentation; 2023. Available from: `https://arxiv.org/abs/2303.09975`.

[18] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images; 2022. Available from: `https://arxiv.org/abs/2201.01266`.

[19] Xing Z, Ye T, Yang Y, Liu G, Zhu L. SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation; 2024. Available from: `https://arxiv.org/abs/2401.13560`.

[20] Danckaers F, Huysmans T, Lacko D, Ledda A, Verwulgent S, Dongen SV, et al. Correspondence Preserving Elastic Surface Registration with Shape Model Prior. In: 2014 22nd International Conference on Pattern Recognition. IEEE; 2014. p. 2143-8.

[21] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research. 2005;30(1):79-82.

[22] Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development. 2014;7(3):1247-50.

[23] Shan Qing Yeoh P, Bing L, Li Goh S, Hasikin K, Wu X, Chai Hum Y, et al. An Efficient Neural Network for Segmenting Multiple Joint Tissues From Knee MRI With Hyperparameter Optimization: Data From the Osteoarthritis Initiative. IEEE Access. 2024;12:123757-70.

[24] Kakavand R, Palizi M, Tahghighi P, Ahmadi R, Gianchandani N, Adeeb S, et al. Integration of Swin UNETR and statistical shape modeling for a semi-automated segmentation of the knee and biomechanical modeling of articular cartilage. Scientific Reports. 2024;14(1):2748.

[25] Van den Broeck J, Vereecke E, Wirix-Speetjens R, Vander Sloten J. Segmentation accuracy of the knee joint using MRI technology. Published for the British Editorial Society of Bone & Joint Surgery by Churchill Livingstone; 2013-01-01.

[26] Starmans MPA, van der Voort SR, Castillo Tovar JM, Veenland JF, Klein S, Niessen WJ. Chapter 18 - Radiomics: Data mining using quantitative medical image features. In: Zhou SK, Rueckert D, Fichtinger G, editors. Handbook of Medical Image Computing and Computer Assisted Intervention. The Elsevier and MICCAI Society Book Series. Academic Press; 2020. p. 429-56.

[27] Zaman FA, Wu X, Xu W, Sonka M, Mudumbai R. Trust, but Verify: Robust Image Segmentation using Deep Learning. In: 2023 57th Asilomar Conference on Signals, Systems, and Computers; 2023. p. 1070-4.

[28] Coleman JL, Widmyer MR, Leddy HA, Utturkar GM, Spritzer CE, Moorman CT, et al. Diurnal variations in articular cartilage thickness and strain in the human knee. Journal of Biomechanics. 2013;46(3):541-7.

[29] Braun HJ, Gold GE. Advanced MRI of articular cartilage. Imaging in medicine. 2011 10;3:541-55.

[30] Dyck PV, Smekens C, Roelant E, Vyvere TV, Snoeckx A, Smet ED. 3D CAIPIRINHA SPACE versus standard 2D TSE for routine knee MRI: a large-scale interchangeability study. European Radiology. 2022 3;32:6456-67.

[31] Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Real-Time Prediction of Segmentation Quality. Cham: Springer International Publishing; 2018.

[32] Valindria VV, Lavdas I, Bai W, Kamnitsas K, Aboagye EO, Rockall AG, et al. Reverse Classification Accuracy: Predicting Segmentation Performance in the Absence of Ground Truth. IEEE Transactions on Medical Imaging. 2017;36(8):1597-606.

[33] Huysentruyt R, Van den Borre I, Lazendić S, Duquesne K, Van Oevelen A, Li J, et al. Sample Size Effect on Musculoskeletal Segmentation: How Low Can We Go? Electronics. 2024;13(10).

[34] Hoebel K, Andrearczyk V, Beers A, Patel J, Chang K, Depeursinge A, et al. An exploration of uncertainty information for segmentation quality assessment. In: Išgum I, Landman BA, editors. Medical Imaging 2020: Image Processing. vol. 11313. International Society for Optics and Photonics. SPIE; 2020. p. 113131K.

[35] Zaman FA, Zhang L, Zhang H, Sonka M, Wu X. Segmentation quality assessment by automated detection of erroneous surface regions in medical images. Computers in Biology and Medicine. 2023;164:107324.

[36] Lu J, Li W, Wang Q, Zhang Y. Research on Data Quality Control of Crowdsourcing Annotation: A Survey. In: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech); 2020. p. 201-8.

Chapter 11

Automated anatomical landmark-based applications

In this chapter, two potential applications of automated landmarking are illustrated: the search for a reliable measurement of femoral torsion and the estimation of meniscus geometry from the bony anatomy. Registered surfaces (see Section 11.3. Surface registration) or a dense set of anatomically corresponding landmarks are fundamental prerequisites for both applications.

11.1 Application: Femoral anteversion measurement

This work was previously published in a more elaborated version as: *Van Fraeyenhove, B., Verhaegen, J. C. F., Grammens, J., Mestach, G., Audenaert, E., Van Haver, A., Verdonk, P. (2023). The quest for optimal femoral torsion angle measurements: a comparative advanced 3D study defining the femoral neck axis. Journal of experimental orthopaedics, 10(1), 141. https://doi.org/10.1186/s40634-023-00679-9*

11.1.1 Introduction

The femoral torsion is measured by the angle between between the femoral neck axis (FNA, proximal femur) and the posterior condyle line (PCL, distal femur) [1]. It is a key factor in hip and knee biomechanics, influencing conditions like hip dysplasia, patellofemoral instability, and anterior cruciate ligament injury [2, 3, 4]. Accurate measurement is essential for both diagnostics and surgical planning [5, 6, 7, 8]. Traditional 2D methods, such as those described by Murphy [9] are widely used but can be limited by the complex 3D structure of the femoral neck, leading to inconsistent results [10, 11]. Recent advancements in 3D imaging techniques are promising to offer improved accuracy. These new methods, however, vary in their approach to defining the femoral neck axis (FNA) [10, 11], particularly when accounting for abnormalities such as CAM deformities, which affect 34% of males and 20% of females [12, 13]. This study aims to compare five 3D femoral torsion measurement methods against the

standard 2D technique, assess the impact of CAM deformities, and raise awareness of differences in measurement techniques.

11.1.2 Materials and methods

Study Design

This study used 3D models from CT scans of 102 dry femur specimens, previously utilized in research by Audenaert et al. [14]. Femoral torsion was measured using one 2D CT method (Murphy's method) and five different 3D methods. The femurs were previously [14] categorized into CAM (31.37%) and non-CAM groups (68.63%) based on a previously validated fully automated 3D detection method.

Measurement Techniques

Femoral torsion was calculated as the angle between the femoral neck axis (FNA) and the posterior condyle line (PCL) of the distal femur. Murphy's method (method 0) used two 2D transverse CT slices. The 3D methods varied in their strategies for defining the femoral neck center:

- Method 1: Elliptical least-square fitting on the femoral neck midsection [14].
- Method 2: Calculation of the center of mass for the entire femoral neck. (Figure 10.1(a))
- Method 3: Center of mass for the most cylindrical part of the femoral neck. (Figure 10.1(b))
- Method 4: Intersection of the femoral neck with a sphere enlarged by 25%. (Figure 10.2(a))
- Method 5: Intersection with a sphere enlarged by 40%.(Figure 10.2(b))

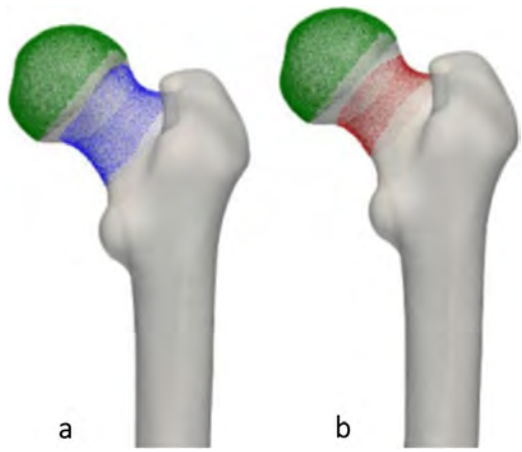


Figure 11.1: Methods 2 and 3 to assess femoral neck using 3D technology.
(a) Method 2: Centre of mass of the complete femoral neck surface.
(b) Method 3: Centre of mass of the most cylindrical part of the femoral neck surface.

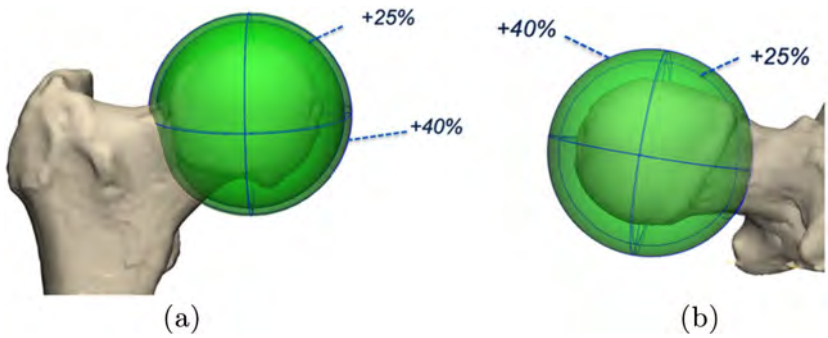


Figure 11.2: Methods 4 and 5 to assess femoral neck using 3D technology. A best fit sphere of the femoral head was drawn. Next, a second sphere was generated by increasing the radius with 25% (**(a)** method 4) and 40% (textbf(b) method 5).

While method 4 and 5 were performed rather manually in the Mimics software package (Materialise NV, Leuven, Belgium) for this study, both methods are perfect candidates for extensive automation based on automated landmark detection and anatomically corresponding landmarks.

Statistical Analysis

Results were presented as mean ± standard deviation, with parametric testing conducted after verifying normal distribution. ANOVA with Bonferroni post hoc testing was used to evaluate differences between methods, and Pearson correlation analysis assessed the strength of associations.

11.1.3 Results

The study compared femoral torsion measurements from one 2D-CT method [9] and five different 3D methods. Measurement statistics are summarized in the box-and-whisker diagrams in Figure 10.3. Key findings include:

- Higher values for 3D methods: The 2D-CT method [9] measured femoral torsion at 8.12° ± 7.30°. In contrast, the 3D methods showed higher mean values, ranging from 8.21° to 13.21°.
- Differences among 3D methods: Method 2 had the highest torsion (13.21° ± 8.60°), and method 3 had the lowest (8.21° ± 7.64°). Statistically significant differences were found between most methods ($p < 0.001$), except between methods 0 and 3, and 1 and 5.
- CAM deformity effects: Hips with a CAM deformity showed a trend of higher torsion measurements, but the differences were not statistically significant. Method 4 might underestimate torsion in the presence of CAM, making method 5 a more reliable option.

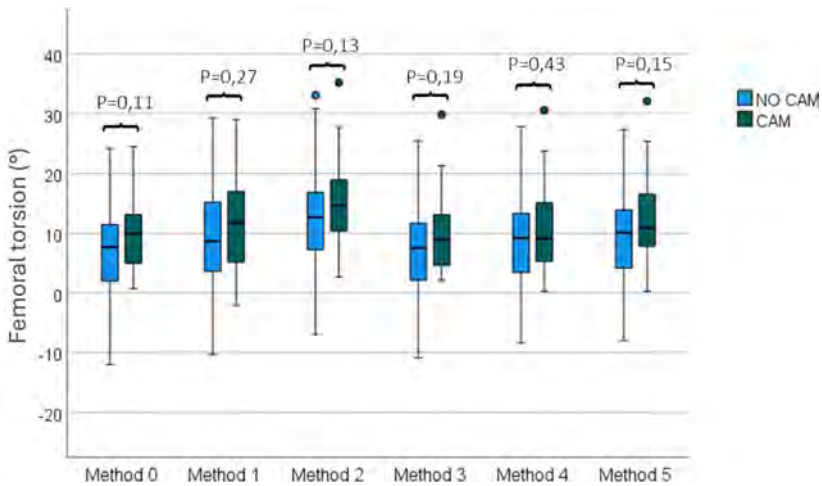


Figure 11.3: Box-and-whisker diagram: Femoral torsion (°), measured with one 2D method (Method 0) and 5 different 3D methods in the NO CAM and CAM group. Independent-samples t-test showed no statistically significant difference between non-CAM and CAM group (P>0.05).

11.1.4 Discussion

This study demonstrates that variations in femoral torsion values are more influenced by the measurement technique than by patient characteristics. The traditional 2D measurement (Murphy's method) differs significantly from most 3D methods, suggesting that 2D techniques may not be as reliable. Femoral torsion tends to be underestimated with 2D methods, especially compared to 3D approaches, which offer

higher accuracy and consistency by eliminating the projection variability inherent in 2D methods.

Despite a general trend toward higher torsion values in femurs with CAM deformities, this difference was not statistically significant. The study emphasizes that clinicians should always interpret expected torsion values relative to the specific measurement technique used and the presence or absence of CAM deformities. Given the significant variation between 2D and 3D measurements (up to 5° on average with method 2), the choice of technique can have critical implications in surgical decision-making, particularly in femoral derotational osteotomy.

Further research involving larger, asymptomatic populations is desirable to validate these findings and optimize the reliability of measurement techniques in clinical practice. Last but not least, investigation of the interaction with other morphological hip parameters, including the pelvic morphology could be of great importance in better understanding the mechanisms behind the observed variability in the femoral torsion.

In conclusion, among the 3D methods, methods 4 and 5 stand out as the most reliable, though method 4 may underestimate torsion in cases with a CAM deformity due to the placement of the intersection point on the femoral neck. Method 5, which is less affected by the presence of CAM deformities, is thus recommended for clinical use in assessing femoral torsion.

11.2 Application: Meniscal Anatomy Estimation

This work is summarized here for illustrative purposes, with a focus on the meniscal anatomy. It was previously published in a more elaborate version as: *Van Oevelen, A., Duquesne, K., Peiffer, M., Grammens, J., Burssens, A., Chevalier, A., Steenackers, G., Victor, J., Audenaert, E. (2023). Personalized statistical modeling of soft tissue structures in the knee. Frontiers in bioengineering and biotechnology, 11, 1055860. https://doi.org/10.3389/fbioe.2023.10558609*

11.2.1 Introduction

Osteoarthritis (OA) affects nearly 25% of the global population and is one of the most rapidly growing socio-economic burdens [15, 16]. Knee OA accounts for 83% of this burden [17]. Despite its prevalence, understanding the interplay between biomechanical and systemic factors in OA progression remains limited [18]. Current methods for accurately measuring in vivo knee joint forces and analyzing soft tissue function are insufficient, hindering research. Computational musculoskeletal modeling offers a non-invasive way to estimate joint mechanics and anatomical variance, but it heavily depends on accurate anatomical data, often obtained through labor-intensive manual segmentation of CT or MRI scans [19, 20]. Manual methods are time-consuming, prone to errors, and face limitations in accessibility and cost, particularly with MRI [21, 22].

A promising alternative is combining musculoskeletal modeling with statistical shape analysis to address these challenges. Audenaert et al. developed a validated pipeline for semi-automated lower limb segmentation from CT data [23, 24]. Van Houcke et al. advanced this by predicting cartilage geometry using MRI datasets, improving efficiency and reducing manual segmentation errors [25]. These methods have been applied successfully to the ankle joint [26], but comparable advances, without the need for extensive image segmentation, are still limited for the knee.

This study aims to develop a scalable computational model for patient-specific knee joint soft tissue anatomy, avoiding manual segmentation. The study's objectives include predicting the cartilage layers and static meniscal geometry, including the validation of this algorithm.

11.2.2 Materials and methods

Data Collection

Two imaging datasets were used: a CT dataset for bone modeling and an MRI dataset for soft tissue features. The CT dataset comprised 311 bilateral lower limb scans (n=622) from 181 males and 130 females, aged 67.8 ± 10.8 and 69 ± 13.3 years, respectively. The MRI dataset involved 53 healthy Caucasian males (age 17–25) who underwent high-resolution scans of the hip, knee, and ankle using a Siemens 3 Tesla MRI.

MRI Segmentation and Feature Identification

MRI data were imported into Mimics software, and osseous and cartilage anatomy was segmented from 53 cases. Osseous structures were modeled using a semi-automated segmentation based on Statistical Shape Models (SSMs), achieving high accuracy for femur and tibia segmentation [24]. Anatomical landmarks for meniscal anatomy (multiple meniscal root attachment points) were identified in 10 cases.

Anatomical correspondence-based landmark transfer

The previously annotated anatomical landmarks were averaged to reduce the meniscal root attachment sites to a single point. Non-rigid surface registration (see also Section 11.3. Surface registration) then enabled the transfer of these points towards the SSM, taking advantage of the established anatomical correspondences. Finally, these anatomical landmarks can now be inferred for any new patient case by fitting the SSM.

Meniscal anatomy prediction

The menisci were modeled following their general geometric characteristics: starting from their tibial attachments and having a triangular radial cross-section (with height and width derived from a polynomial fitting in function of the relative meniscus length), thereby circumferentially wrapping around the space between the tibial plateau and the femoral condyles (Fig. 10.4).

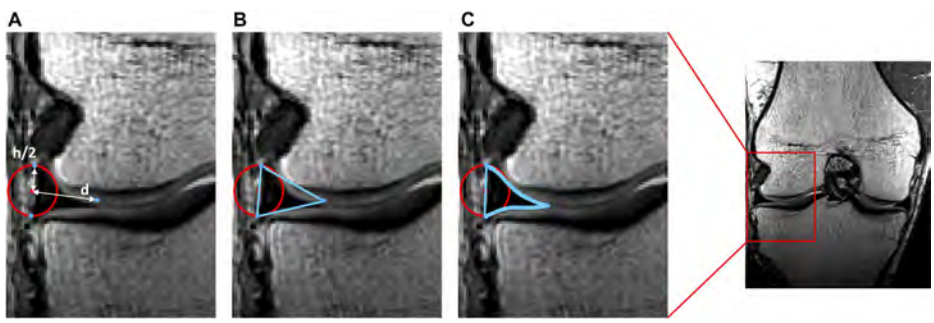


Figure 11.4: (A) A tube with the same diameter as the estimated meniscal height (external rim) was wrapped in a circumferential way from the anterior to the posterior horn. A varying distance (d) is employed for every cross-section. (B) Triangles are formed by interconnecting the three points per radial cross-section. (C) Following correction for local penetration, the edges of the meniscus adapt to fit in between cartilage layers.

Validation

Validation was conducted through leave-one-out experiments, comparing model inferred and manually segmented anatomy using Root Mean Square Error (RMSE), Average Surface Distance (ASD), and Hausdorff Distance (HD). Validation covered meniscal root attachment positions, and meniscal anatomy predictions (Fig. 10.5) for 10 previously unseen cases by the model.

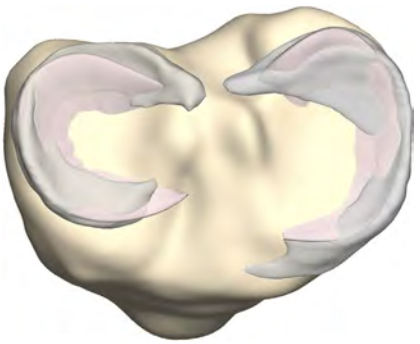


Figure 11.5: Axial view on the tibial plateau (yellow). The manually segmented medial and lateral menisci (grey) serve as ground truth to compare the superimposed inferred medial and lateral menisci (pink).

11.2.3 Results

Validation of meniscal root attachment identification

Following RMSE were observed: for the medial meniscus anterior root 3.51 mm, for the medial meniscus posterior root 2.71 mm, for the lateral meniscus anterior root 3.75 mm and for the lateral meniscus posterior root 2.82 mm. Errors were in the range between 0 (lateral meniscus posterior root) and 5.29 mm (lateral meniscus anterior root).

Validation of meniscal geometry prediction

Accuracy was comparable for medial and lateral meniscus anatomy prediction, with a median RMSE of 2.93 mm (medial, range 1.85 - 4.66 mm) and 2.04 mm (lateral, range 1.88 - 3.29 mm). An average point-dependent error between 0 and 5 mm was observed (Figure 10.6). The largest errors were observed at the inner rim of the lateral meniscus and the anterior root of the medial meniscus (red on the color map in Figure 10.6).

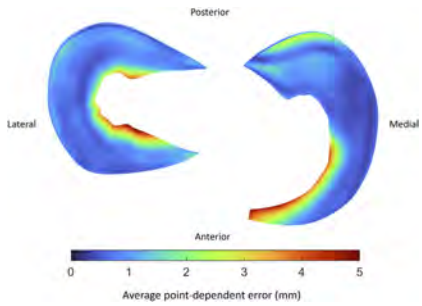


Figure 11.6: Average point-dependent error between inferred and manually segmented menisci, color mapped on the mean lateral (left) and mean medial (right) meniscus.

11.2.4 Discussion

In this study, meniscal anatomy was modeled as mobile and elastic, adapting to the variable shapes of the femoral condyles and tibial positions. The prediction of meniscal geometry was based on the underlying osseous structure. The validation study (compared with manual segmentation) reported reasonable root mean square errors (RMSE) for inter- and intra-observer reliability during manual segmentation, aligning with findings from prior studies [27, 28, 29].

A significant challenge identified in the prediction of meniscal anatomy was the variation in the anterior root morphology of the medial meniscus. Specifically, one case displayed an extreme anterior insertion, contributing to a larger prediction error. This observed anatomical variation was consistent with earlier studies [30], with a reported prevalence of approximately 7% of the population. To improve future models, probabilistic approaches could be incorporated to account for these distinct types of anterior medial meniscal roots.

The overall errors in meniscal geometry prediction were comparable to those observed in previous studies [27, 28], with the largest errors noted for the anterior root of the medial meniscus and the inner rim of the lateral meniscus. The modeling method used polynomial functions to describe meniscal geometry, enabling to vary the height and width over the course of the meniscus, with local adjustments made to prevent penetration into adjacent cartilage and bone structures.

In conclusion, this study validates a method to accurately estimate the meniscal anatomy automatically, without the need for time-consuming manual segmentation. The model showed a level of accuracy, comparable to manual soft tissue segmentation. It offered a solid foundation for an extension towards dynamic meniscus modeling in weight-bearing conditions and over the full range of motion of the knee [31].

Bibliography

[1] Høiseth A, Reikerås O, Fønstelien E. Evaluation of Three Methods for Measurement of Femoral Neck Anteversion. Acta Radiologica. 1989 1;30:69-73.

[2] Lerch TD, Todorski IAS, Steppacher SD, Schmaranzer F, Werlen SF, Siebenrock KA, et al. Prevalence of Femoral and Acetabular Version Abnormalities in Patients With Symptomatic Hip Disease: A Controlled Study of 538 Hips. The American Journal of Sports Medicine. 2018 1;46:122-34.

[3] Lee TQ, Morris G, Csintalan RP. The Influence of Tibial and Femoral Rotation on Patellofemoral Contact Area and Pressure. Journal of Orthopaedic & Sports Physical Therapy. 2003 11;33:686-93.

[4] Alpay Y, Ezici A, Kurk MB, Ozyalvac ON, Akpinar E, Bayhan AI. Increased femoral anteversion related to infratrochanteric femoral torsion is associated with ACL rupture. Knee Surgery, Sports Traumatology, Arthroscopy. 2020 8;28:2567-71.

[5] Fabricant PD, Bedi A, Torre KDL, Kelly BT. Clinical Outcomes After Arthroscopic Psoas Lengthening: The Effect of Femoral Version. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2012 7;28:965-71.

[6] Fabricant PD, Fields KG, Taylor SA, Magennis E, Bedi A, Kelly BT. The Effect of Femoral and Acetabular Version on Clinical Outcomes After Arthroscopic Femoroacetabular Impingement Surgery. Journal of Bone and Joint Surgery. 2015 4;97:537-43.

[7] Knafo Y, Houfani F, Zaharia B, Egrise F, Clerc-Urmès I, Mainard D. Value of 3D Preoperative Planning for Primary Total Hip Arthroplasty Based on Biplanar Weightbearing Radiographs. BioMed Research International. 2019 3;2019:1-7.

[8] Moralidou M, Laura AD, Henckel J, Hothi H, Hart AJ. Three-dimensional preoperative planning of primary hip arthroplasty: a systematic literature review. EFORT Open Reviews. 2020 12;5:845-55.

[9] Murphy SB, Simon SR, Kijewski PK, Wilkinson RH, Griscom NT. Femoral anteversion. The Journal of Bone & Joint Surgery. 1987 10;69:1169-76.

[10] Kaiser P, Attal R, Kammerer M, Thauerer M, Hamberger L, Mayr R, et al. Significant differences in femoral torsion values depending on the CT measurement technique. Archives of Orthopaedic and Trauma Surgery. 2016 9;136:1259-64.

[11] Scorcelletti M, Reeves ND, Rittweger J, Ireland A. Femoral anteversion: significance and measurement. Journal of Anatomy. 2020 11;237:811-26.

[12] Hanzlik S, Riff AJ, Wuerz TH, Abdulian M, Gurin D, Nho SJ, et al. The Prevalence of Cam Morphology: A Cross-Sectional Evaluation of 3,558 Cadaveric Femora. Frontiers in Surgery. 2021 1;7:588535.

[13] Morales-Avalos R, Tapia-Náñez A, Simental-Mendía M, Elizondo-Riojas G, Morcos-Sandino M, Tey-Pons M, et al. Prevalence of Morphological Variations Associated With Femoroacetabular Impingement According to Age and Sex: A Study of 1878 Asymptomatic Hips in Nonprofessional Athletes. Orthopaedic Journal of Sports Medicine. 2021 2;9:232596712097789.

[14] Audenaert EA, Baelde N, Huysse W, Vigneron L, Pattyn C. Development of a three-dimensional detection method of cam deformities in femoroacetabular impingement. Skeletal Radiology. 2011 7;40:921-7.

[15] Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, de Almeida RC, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021 11;184:6003-5.

[16] Hunter DJ, March L, Chew M. Osteoarthritis in 2020 and beyond: a Lancet Commission. The Lancet. 2020 11;396:1711-2.

[17] Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012 12;380:2163-96.

[18] Sharma L, Song J, Dunlop D, Felson D, Lewis CE, Segal N, et al. Varus and valgus alignment and incident and progressive knee osteoarthritis. Annals of the Rheumatic Diseases. 2010 11;69:1940-5.

[19] Marra MA, Vanheule V, Fluit R, Koopman BHFJM, Rasmussen J, Verdonschot N, et al. A Subject-Specific Musculoskeletal Modeling Framework to Predict In Vivo Mechanics of Total Knee Arthroplasty. Journal of Biomechanical Engineering. 2015 2;137.

[20] Kang KT, Kim SH, Son J, Lee YH, Koh YG. Validation of a computational knee joint model using an alignment method for the knee laxity test and computed tomography. Bio-Medical Materials and Engineering. 2017 7;28:417-29.

[21] Seim H, Kainmueller D, Heller M, Lamecker H, Zachow S, Hege HC. Automatic segmentation of the pelvic bones from CT data based on a statistical shape model. In: Proceedings of the First Eurographics Conference on Visual Computing for Biomedicine. EG VCBM'08. Goslar, DEU: Eurographics Association; 2008. p. 93–100.

[22] Bae KT, Shim H, Tao C, Chang S, Wang JH, Boudreau R, et al. Intra- and inter-observer reproducibility of volume measurement of knee cartilage segmented from the OAI MR image set using a novel semi-automated segmentation method. Osteoarthritis and Cartilage. 2009 12;17:1589-97.

[23] Haver AV, Mahieu P, Claessens T, Li H, Pattyn C, Verdonk P, et al. A statistical shape model of trochlear dysplasia of the knee. The Knee. 2014 3;21:518-23.

[24] Audenaert EA, Houcke JV, Almeida DF, Paelinck L, Peiffer M, Steenackers G, et al. Cascaded statistical shape model based segmentation of the full lower limb in CT. Computer Methods in Biomechanics and Biomedical Engineering. 2019 4;22:644-57.

[25] Houcke JV, Audenaert EA, Atkins PR, Anderson AE. A Combined Geometric Morphometric and Discrete Element Modeling Approach for Hip Cartilage Contact Mechanics. Frontiers in Bioengineering and Biotechnology. 2020 4;8.

[26] Peiffer M, Burssens A, Duquesne K, Last M, Mits SD, Victor J, et al. Personalised statistical modelling of soft tissue structures in the ankle. Computer Methods and Programs in Biomedicine. 2022 5;218:106701.

[27] Vrancken ACT, Crijns SPM, Ploegmakers MJM, O'Kane C, van Tienen TG, Janssen D, et al. 3D geometry analysis of the medial meniscus – a statistical shape modeling approach. Journal of Anatomy. 2014 10;225:395-402.

[28] Tack A, Mukhopadhyay A, Zachow S. Knee menisci segmentation using convolutional neural networks: data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage. 2018 5;26:680-8.

[29] Gao KT, Xie E, Chen V, Iriondo C, Calivà F, Souza RB, et al. Large-Scale Analysis of Meniscus Morphology as Risk Factor for Knee Osteoarthritis. Arthritis & Rheumatology. 2023 11;75:1958-68.

[30] Coninck TD, Vanrietvelde F, Seynaeve P, Verdonk P, Verstraete K. MR imaging of the anatomy of the anterior horn of the medial meniscus. Acta Radiologica. 2017 4;58:464-71.

[31] Oevelen AV, Peiffer M, Chevalier A, Victor J, Steenackers G, Audenaert E, et al. The relation between meniscal dynamics and tibiofemoral kinematics. Scientific Reports. 2024 4;14:8829.

Chapter 12

Small medial femoral condyle: a pilot study

This pilot study originates from an empirical observation in clinics, which could not be explained by the current scientific evidence. Multiple risk factors for early onset and rapid progression of knee OA are already known (see 2.4.1: Epidemiology and pathophysiology). Yet, for a subpopulation of medial compartment degeneration patients at a young age, none of the known risk factors were present. On cross-sectional views of the MRI scan however, a rather narrow medial femoral condyle was observed. In this pilot study, the morphology of distal femur and proximal tibia was compared between a group of suspected small medial femoral condyle morphotype patients and a control group (imaging for minor acute trauma). Both a landmark-based method (see 9: Automated Landmark Annotation for Morphometric Analysis of Distal Femur and Proximal Tibia) and statistical parametric mapping (see chapter7: Statistical parametric mapping for segmentation evaluation) were applied to characterize the small medial femoral condyle morphotype and assess the link with early medial compartment degeneration.

This work was previously published as: *Grammens, J., Van Haver, A., Danckaers, F., Booth, B., Sijbers, J., Verdonk, P. (2021). Small medial femoral condyle morphotype is associated with medial compartment degeneration and distinct morphological characteristics: a comparative pilot study. Knee Surgery, Sports Traumatology, Arthroscopy, 29(6), 1777–1789. https://doi.org/10.1007/s00167-020-06218-8*

12.1 Introduction

The relationship between specific knee morphotypes and pathology has already been investigated for several morphological variations [1, 2, 3, 4, 5, 6, 7]. Implications posed by these distinct morphotypes are of great clinical importance since they may contribute in prevention, lead to a better treatment choice (optimised for each specific patient) and even to new personalised therapies. Based on several morphological studies, it can be concluded that the standard treatment may not meet the needs of certain groups of patients. Furthermore, the huge variability of the coronal alignment, reported in both

osteoarthritic and non-osteoarthritic knees suggests a more individualized treatment approach in restoring the functionality of the knee [8, 9, 10]. In addition to the coronal joint line alignment phenotype, the morphotype of the knee joint (shape of the distal femur and proximal tibia) is a complementary concept to describe and investigate anatomical variations in relation to early degeneration of the knee joint; smaller joint contact surfaces may increase the contact stress and might lead to overload.

Though increased shape-related stresses are difficult to investigate in living patients, specific knee morphotypes are already known to be related to certain pathologies. Trochlear dysplasia is characterized by a reduced contact surface area in the patellofemoral joint and is associated with patellar instability and early patellofemoral osteoarthritis [11, 12, 13]. Lateral femoral condylar hypoplasia is associated with a valgus alignment and lateral knee osteoarthritis [6]. In addition to the already documented associations between shape and pathology, a specific knee morphotype, characterized by a smaller medial femoral condyle (SMC), may also result in an increased risk for degenerative changes in the medial compartment of the knee. This morphotype has not been described before and this study might be the first step in exploring the link between bone morphology and early degeneration in the medial compartment. The aim of this study is to identify knee joint shape differences between SMC knees and a control group and to assess the presence of medial compartment degeneration in both groups. The study hypothesis is that the distal femur and proximal tibia bone shapes of the SMC group differ from those of a control group and that SMC knees demonstrate higher incidence of medial compartment degeneration. To test this hypothesis two complementary approaches were applied; a landmark-based shape analysis to evaluate a set of pre-defined parameters and a global shape analysis to evaluate local differences. By describing the small medial femoral condyle as a new knee morphotype, this concept can then play an important clinical role in the treatment and the understanding of the multifactorial process of early-onset osteoarthritis (OA) in post-meniscectomised knees.

12.2 Materials and methods

As this is a retrospective study, IRB approval was obtained from the local ethical committee (AZ Monica OG106, study ID number 413) and all persons gave their informed consent prior to their inclusion in the study. Similar to previous three-dimensional (3D) morphometric studies, a validated landmark-based 3D analysis was performed on the distal femur and proximal tibia to assess the knee joint geometry, including the cartilage [14, 15]. A set of predefined landmarks was manually identified and used to scale the knees isotropically to match the standard size. Additionally, a set of reference planes was also constructed based on these landmarks. This method allows quantification of a predefined set of parameters, which may reveal morphometric differences between control group and SMC knees which have (to our knowledge) not been reported before. Furthermore, a more innovative global shape analysis was performed on the complete bone shape surfaces using statistical parametric mapping. This method is not limited to a number of predefined validated landmarks but includes all points of the bone surfaces and may reveal differences that were not captured in the discrete landmark-based analysis. A flowchart of the study design is presented in Figure 8.1.

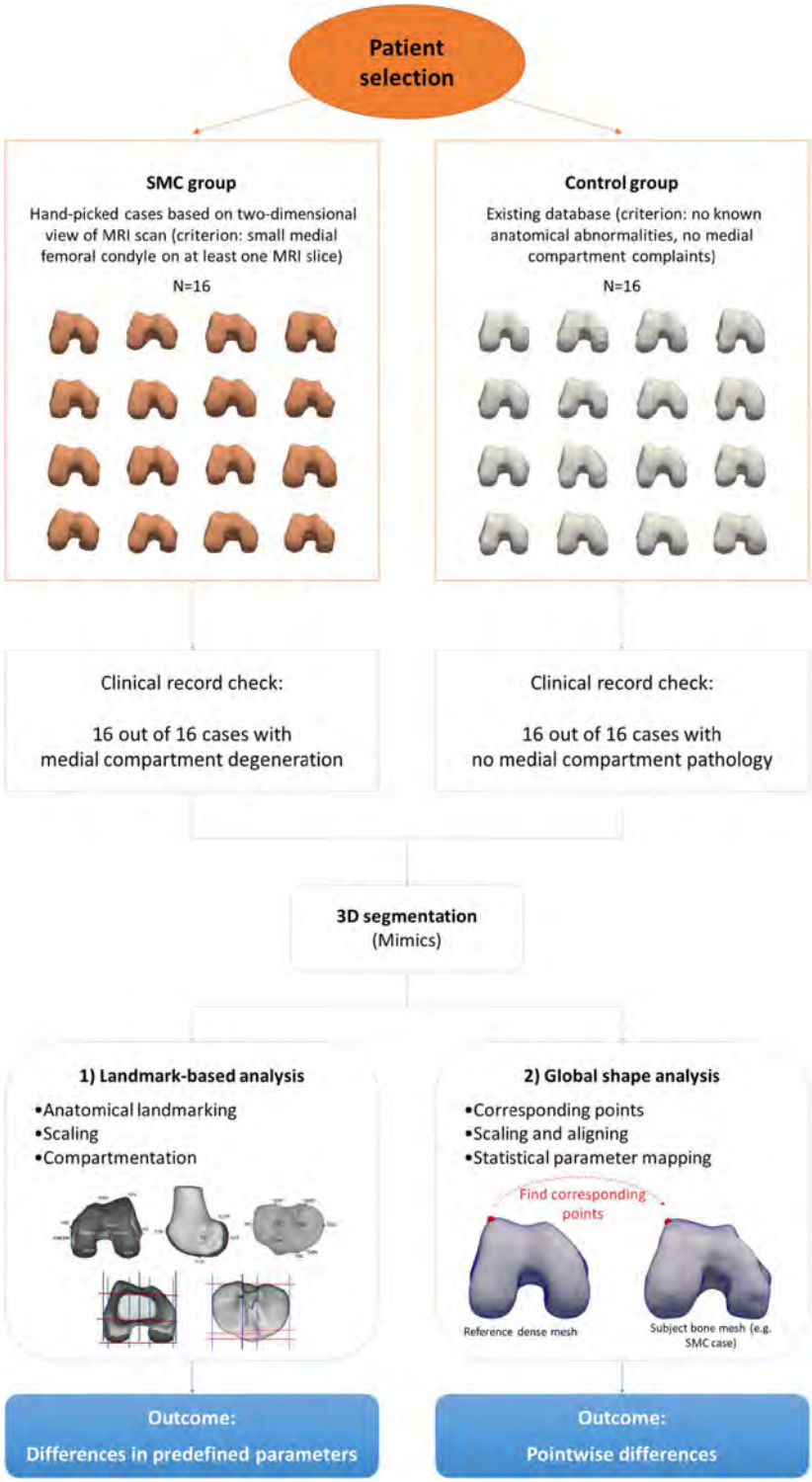


Figure 12.1: Methodology flowchart. A database of medical imaging was constructed for two groups: the small medial condyle group and a control group. 3D models from tibia and femur were constructed from the medical imaging data and post-processed in a landmark-based analysis and a global shape analysis.

12.2.1 Study population

The SMC group included 16 patients (8 females and 8 males; mean age 39 +/- 14 years) characterized by a small medial compartment. These patients were selected from a database which was built between 2015 and 2019 by the senior author, who included patients with distinct knee joint anatomy. The selection of the 16 SMC patients was primarily based on the presence of a small medial femoral condyle, observed on at least one slice of the MRI. Secondly, patients with a short bone stock (caused by limited region of interest) on the MRI and patients who underwent knee surgery prior to the MRI were excluded from this study. As a control group, 16 patients without anatomic knee abnormalities were selected (8 females and 8 males; mean age 30 +/- 9 years). The control group consulted the orthopaedic surgeon for a minor acute trauma and served already as control group in several other studies [14, 15, 16]. Imaging was performed by means of a CT arthrography. After selecting the patients based on their morphology and the quality of the MRI, the medical history and medical images of the SMC and the control group were inspected in detail to assess medial compartment degeneration.

12.2.2 Generation and isotropic scaling of 3D computer models

The medical imaging (MRI or CT) was performed in supine position with 0° of knee flexion and the toes pointing straight up. The images were loaded in a 3D image processing software system (Mimics 22.0, Materialise, Haasrode, Belgium) to create 3D models of the knee bones including the cartilage. Differences in knee size may significantly affect the metric measurements. To avoid this, the 3D models were isotropically resized before the analyses to exclude size differences. A generalized Procrustes transformation of the surface models was applied, using a custom developed code in MATLAB (Matlab 9.6.0, R2019a, Mathworks, Natick, MA), to minimize the pose and size variance between the knees while preserving the shape and underlying proportions of the knees [17]. The applied Procrustes transformation consists of a combination of translation, rotation and isotropic (same amount in the three dimensions) scaling. The standard size was determined by calculating the average femoral shape of the control group. Twelve validated anatomical landmarks, covering the extremes of the distal femur in the anteroposterior (AP), mediolateral (ML) and proximodistal (PD) direction, were defined to calculate a rescaling factor for each knee. This factor was considered as a measure of the femoral size and was also used to isotropically resize the respective tibiae.

12.2.3 Definition of the landmarks

Anatomical landmarks on the femur and tibia were defined in Mimics (Fig. 8.1). Both the 3D models and medical images were used to ensure a precise location of the landmarks. The landmarks are described in Appendix A and are identical to those described in our previous studies [14, 15] Therefore, the ICC of 0.99 and mean error of 1.0 +/- 1.5 mm previously reported for this landmarking technique are also applicable to the landmark positions in this study. In addition to the twelve anatomical landmarks which were used to rescale the bone models, eleven other landmarks on the femur and nine landmarks on

the tibia were defined to create the reference planes and to measure the morphometric characteristics [14, 15].

12.2.4 Definition of the reference planes

The reference planes are predefined by the authors in Mimics in the ML, AP and PD direction. An overview table can be found in the "Appendix" of this article. By identifying the landmarks, the reference planes are automatically fitted on the geometry of the distal femur and proximal tibia. All planes but four are identical to the ones previously described in our other studies [14, 15]. The newly introduced planes are related to the ML width of the medial and lateral condyle.

12.2.5 Measurements

Based on this set of landmarks and reference planes, 19 morphometric measurements of the 3D models were evaluated as described in the "Appendix" and as summarized in the next paragraph. For the medial and lateral condyle of the femur, the overall AP size and the AP size of the posterior parts were measured separately. In the ML direction, the total width of the femur, the width of the medial condyle, the lateral condyle, and the notch were measured. The tibia measurements were performed in a similar way. The AP size of the total tibial plateau, the medial and lateral tibial plateau were measured. In the ML direction, the width of the complete tibial plateau, the medial and lateral tibial plateaus separately and the intercondylar eminence width were measured. Finally, the medial and lateral tibial spine height was measured in the PD direction. The exact definition of the 19 used sizes and distances can be found in the "Appendix" of this article and is the same as described previously [14, 15].

12.2.6 Statistical analysis

All data analyses for the landmark based analysis and medial compartment degeneration data were performed using IBM SPSS Statistics for Windows (Version 24.0, IBM Corp., Armonk, NY). To evaluate the presence of medial compartment degeneration in the SMC versus the control group, a X2 goodness-of-fit test was performed. The rescaling factor and the morphometric measurements of the control group and the SMC group are reported as median and range. A Mann–Whitney U Test was used to compare the morphological measurements in the SMC group with the data in the control group. To facilitate interpretation of the parameters, the mean results of the SMC group are also converted to a percentage with respect to the mean results of the control group. For all statistical tests, a p-value of less than 0.05 was defined as statistically significant. Sample size calculation for the Mann-Whitney-U test was performed in G*Power (version 3.1.9.7, Universität Kiel, Germany) [18] based on an effect size of 1.0, a significance level alpha of 0.05 and power of 0.85. The resulting minimal sample size for equally sized groups was 16 subjects per group.

12.2.7 Global shape analysis

The global shape analysis aims at finding local shape differences, based on the location of each surface point on the femur and tibia. This analysis consisted of three steps: 1) Find the corresponding points (for every surface, the coordinates that correspond with the same anatomical location are determined) between the individual 3D bone models and a reference 3D bone model of femur and tibia. This process is called surface registration and was performed by implementing the iterative process described by Danckaers et al [19] (see also Section 11.3. Surface registration). This process was performed twice: first with an arbitrary chosen reference bone shape, the second time with the mean bone shape (constructed by calculating the mean coordinates for all corresponding points) as a reference. 2) Each registered femur and tibia were separately mirrored (left knee shapes were mirrored to match right knee shapes), rotated, shifted and isotropically rescaled to the reference femur or tibia shape, based on the correspondences of all surface points that was found in step 1. This is done by a generalized Procrustes analysis, which has the effect of minimizing the distance between all corresponding surface points. 3) Compare each individual femur and tibia 3D model to the reference bone shape by calculating the pointwise distances between them. The mean bone shape of the control group served as the reference bone shape. The perpendicular distances from the reference shape were compared between the groups and shown in a color map, plotted on the reference shape (Fig. 8.3 and 8.4, upper part). If there are no local shape differences, the distances will be equal to zero (indicated in green on the color map). If the SMC bone models are on average smaller at that specific location, distances will be negative (indicated in blue on the color map). Conversely, if the SMC bone model was on average larger, the distances will be positive (and indicated in red on the color map). Statistical analysis of these distances was performed to detect differences between the SMC group and the control group. A permutation student t-test was done with correction for multiple comparisons (False Discovery Rate, FDR) to avoid type I errors due to the multiple tests [20]. The statistical significance level was defined at a q-value of less than 0.05. The significant FDR q-values were then mapped in red on the reference bone shape (Fig. 8.3 and 8.4, lower part).

12.3 Results

12.3.1 Medial compartment degeneration

All SMC patients and none of the control group patients showed medial compartment degeneration. The association between knee type and the presence of medial compartment degeneration was significant, $\chi^2 = 32$ ($p < 0.001$). Hence, small medial condyle knees are more likely to develop medial compartment degeneration.

12.3.2 Landmark-based analysis

The knees in the control group (N = 16) were rescaled with a median rescaling factor of 0.98 (min. 0.86; max 1.08), and the knees in the SMC group (N = 16) with a median rescaling factor of 0.94 (min. 0.80; max 1.06). This difference was statistically significant (p = 0.039). The difference between the smallest (0.80) and largest (1.08) rescaling factor in the total study population (N = 32) was 35%. The morphometric measurements for both femur and tibia are summarized in Table 8.1 and Fig. 8.2. The main findings were mediolaterally a smaller medial femoral condyle, a wider lateral femoral condyle, and a wider total distal femur on a smaller tibial plateau. Additionally, both the medial and lateral femoral condyle where anteroposteriorly smaller, as well as the medial tibial plateau and total tibial plateau.

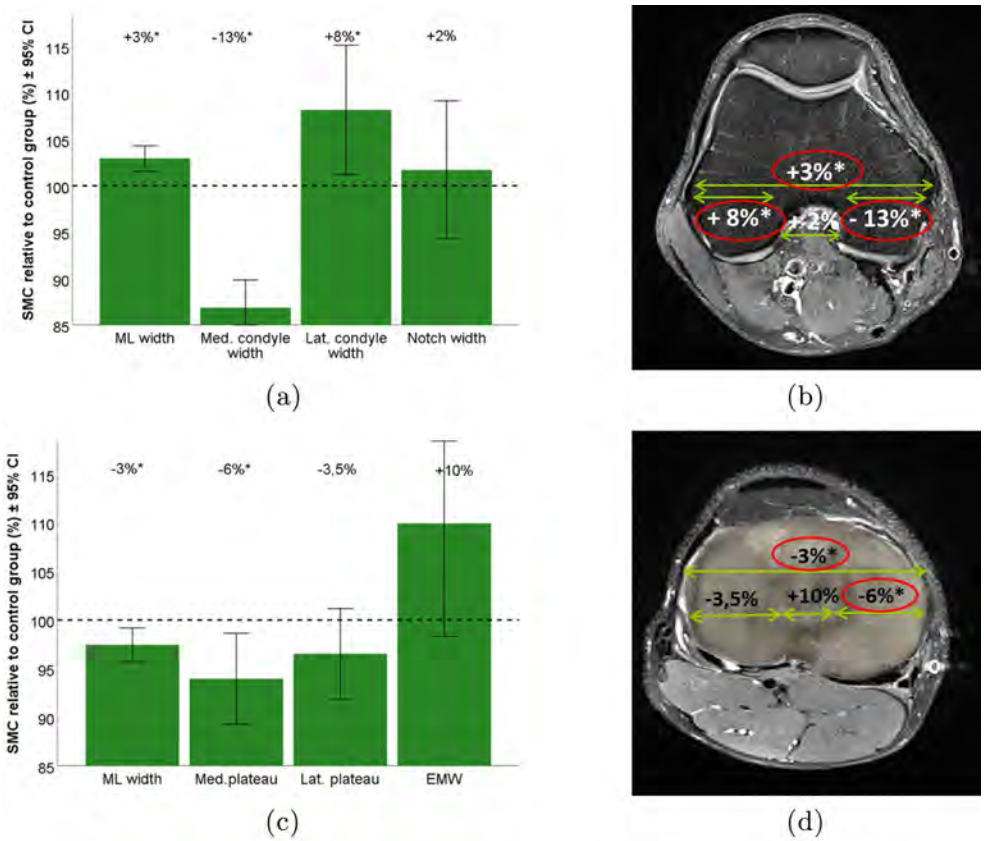


Figure 12.2: (a) and **(c)**: Graph of the mean SMC group ML measurements in relation to the mean control group measurements (dashed line) for femur (a) and tibia (c). **(b)** and **(d)**: Visualization the mean differences for femur (b) and tibia (d) on an axial cross-section of a random subject MRI.

Table 12.1: Summary of the morphometric measurements for the femur and the tibia
n.s.: not statistical significant with a P value higher than .05.

Femur morphometric parameters [mm]	**Median control (range)**	**Median SMC (range)**	***p*-value**
AP depth medial condyle	64 (60–67)	62 (58–65)	0.001
AP depth medial posterior condyle	28 (23–30)	27 (20–33)	n.s.
AP depth lateral condyle	66 (63–69)	65 (61–68)	0.023
AP depth lateral posterior condyle	25 (23–29)	24 (20–29)	n.s.
ML width distal femur	78 (76–82)	82 (77–85)	0.003
ML width medial condyle	24 (21–29)	21 (18–23)	<0.001
ML width lateral condyle	25 (23–29)	28 (21–35)	0.023
ML width intercondylar notch	21 (16–26)	21 (15–27)	n.s.
Tibia morphometric parameters [mm]	**Median control (range)**	**Median SMC (range)**	***p*-value**
AP depth tibial plateau	52 (49–58)	50 (44–53)	0.001
AP depth medial tibial plateau	46 (43–49)	43 (41–49)	0.026
AP position medial tibial spine	20 (15–26)	22 (18–26)	n.s.
AP depth lateral tibial plateau	36 (31–43)	39 (34–42)	n.s.
AP position lateral tibial spine	18 (15–23)	18 (14–24)	n.s.
ML width tibial plateau	76 (71–80)	74 (70–78)	0.026
ML width medial tibial plateau	31 (28–34)	29 (27–36)	0.007
ML width lateral tibial plateau	33 (29–37)	31 (27–40)	n.s.
ML width intercondylar eminence	11 (9–15)	12 (8–16)	n.s.
PD height medial tibial spine	8 (6–10)	9 (7–13)	0.019
PD height lateral tibial spine	7 (4–9)	7 (4–9)	n.s.

12.3.3 Global shape analysis

The results of the global shape analysis of the femur are color-plotted in Fig. 8.3 and all observed trends confirm the results from the landmark-based analysis. The most important significant difference is situated at the inner side of the medial condyle, where the SMC femur is on average 1.1 mm smaller (average FDR q= 0.05) with respect to the mean control femur. The SMC knee showed no statistically significant larger or protruding regions. For the tibia, no significant results were detected (Fig. 8.4). The extreme case shown in fig. 8.5, shows that the combined difference of inner (1.7 mm smaller) and outer side (2.3 mm smaller) of the medial condyle can go up to 4.0 mm (-17% of the average medial condyle width in the control group). The tibia from that same case (Fig. 8.6) showed a 2.1 mm higher lateral spine (+26% of average medial spine height in control group) and 3.5 mm higher medial spine (+50% of average lateral spine height in control group).

12.4 Discussion

The most important finding of this study is the evidence of distinct morphological differences between the control group knee and the small medial condyle knee, which was demonstrated by applying a validated semi-automated landmark-based analysis and a complementary global shape analysis. In the medial compartment, there was a smaller femoral condyle AP and ML, complemented with a smaller tibial plateau ML. However, morphological differences were not limited to the medial compartment, but were also observed at the lateral side. Another remarkable finding was that the overall distal femur was on average wider, while the overall tibial plateau was smaller in the SMC group, which could be an indication for a size and/or shape mismatch between the femoral and tibial geometry. The subjects included in the SMC group were selected by a senior orthopaedic surgeon, based on observations on the MRI images. The patients were not selected based on their medical records, yet it was observed that all patients consulted the surgeon for complaints related to the medial compartment degenerative tearing. As a result of the χ^2-test, the SMC knees in this study have a higher risk to develop OA. Though this study did not investigate a causal relation between shape and pathology, nor the incidence of this shape variation, this study might be the first step in exploring the link between bone morphology and early degeneration in the medial compartment. The selection of the SMC cases was performed based on a two-dimensional (2D) MRI slices, hence dependent on the knee position during image acquisition. In fact, this dependence indicates the need for more advanced 3D like those used in this paper. Although the experience based selection was performed on 2D images, the 3D shape analysis results showed the expected shape difference with respect to the ML width of the medial femoral condyle. As this is an exploratory study, the findings can be used in the future to select knee shapes based on a 3D reality rather than a subjective interpretation of the 2D reconstructions. The underlying mechanism behind the relation between specific morphotypes and pathology has already been investigated for other morphological variations, such as trochlear dysplasia, small notch width and tibia slope. Disturbed biomechanics [1, 2, 3, 4] induce supraphysiological peak stresses and strains [7], which promote early degeneration [21]. A better understanding of the biomechanical implications posed by

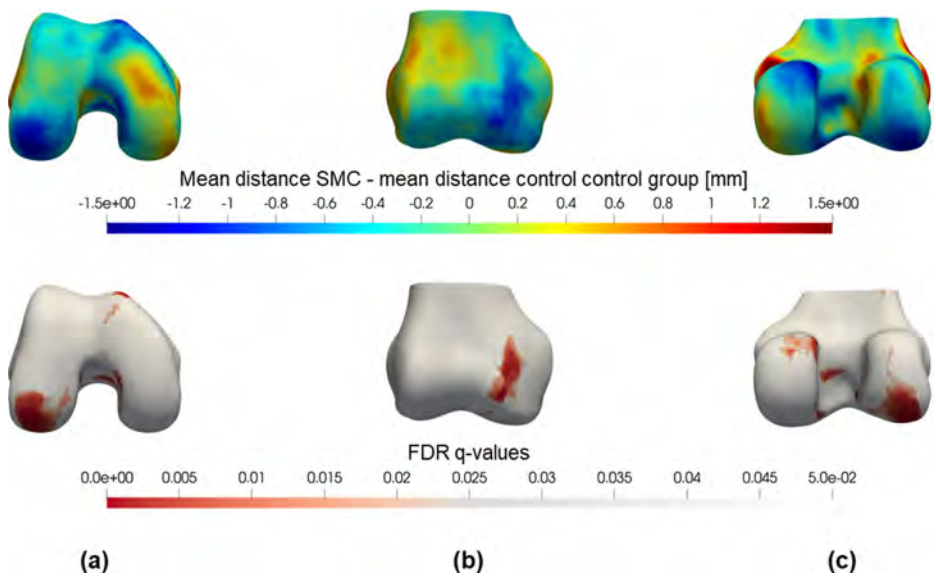


Figure 12.3: (Upper part) color map of the mean femoral shape differences between the SMC and control group. The difference between the mean distances is plotted on the reference femur 3D model (right knee). Blue regions indicate that the SMC femur is on average smaller at that specific location. Conversely, if the SMC femur was on average larger (e.g. caused by a bump or protrusion) the region is colored red. **(Lower part)** FDR q-values for a permutation t-test on femoral shape differences. Distances to the reference femur are compared between the two groups by means of a permutation t-test with 1000 permutations and corrected for false discoveries. Only q-values < 0.05 were plotted in red on the reference femur (right knee). **(a)** distal view; **(b)** anterior view; **(c)** posterior view

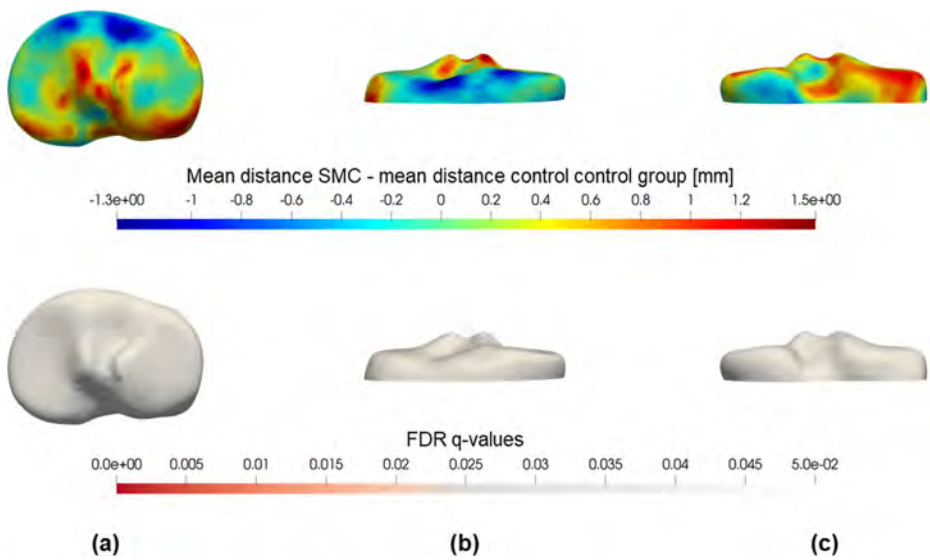


Figure 12.4: (Upper part) Color map of the mean tibial shape differences between the SMC and control group. The difference between the mean distances is plotted on the reference tibia 3D model. Blue regions indicate that the SMC tibia is on average smaller at that specific location. Conversely, if the SMC tibia was on average larger (e.g. caused by a bump or protrusion) the region is colored red. **(Lower part)** FDR q-values for a permutation t-test on tibial shape differences. Distances to the reference tibia are compared between the two groups by means of a permutation t-test with 1000 permutations and corrected for false discoveries. Only q-values < 0.05 were plotted in red on the reference tibia. **(a)** proximal view; **(b)** anterior view; **(c)** posterior view

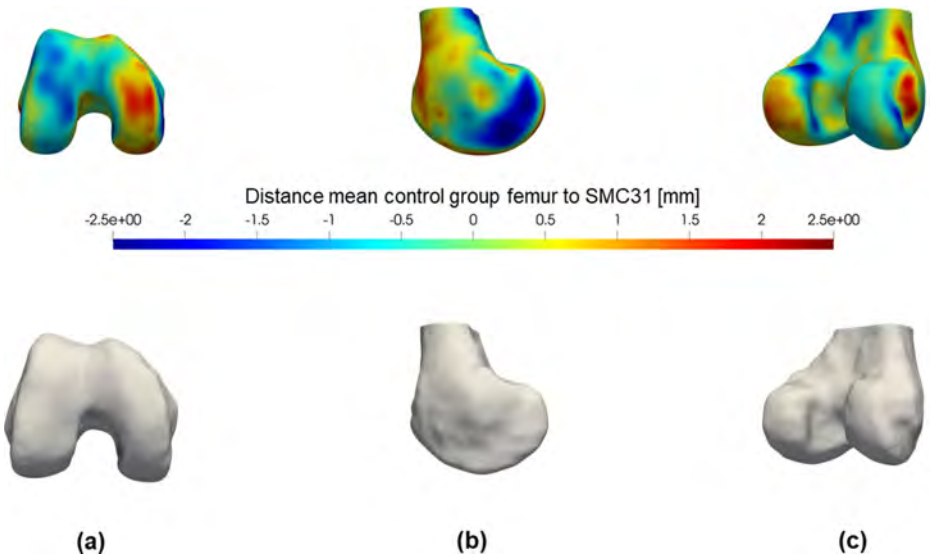


Figure 12.5: (Upper part) Color map of the differences between a right SMC femur and the mean control femur. The distances are plotted on the reference femur 3D model. Blue regions indicate that the SMC femur is smaller at that specific location. Conversely, if the SMC femur was larger (e.g. caused by a bump or protrusion) the region is colored red. **(Lower part)** Example of an SMC femur case. Visualisation of the case used to calculate the distances from the reference femur in the upper part of this image. **(a)** proximal view; **(b)** anterior view; **(c)** posterior view

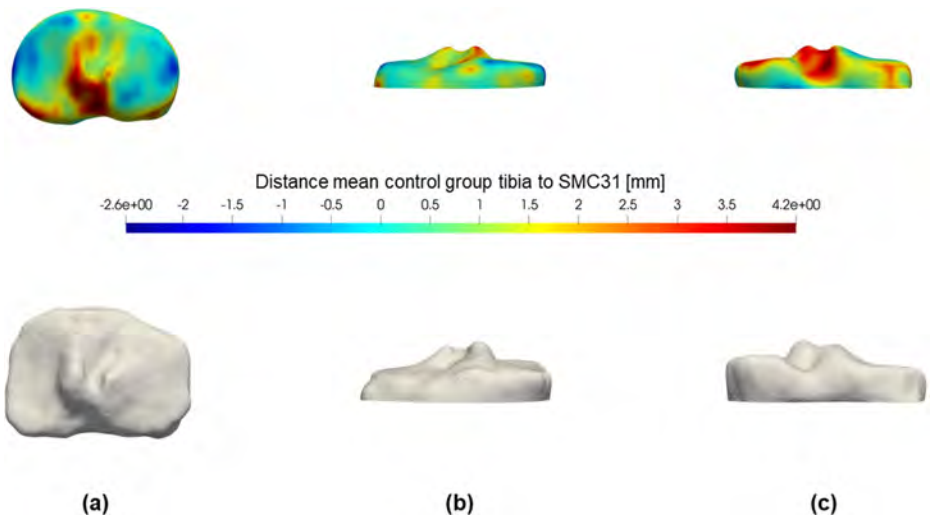


Figure 12.6: (Upper part) Color map of the differences between a right SMC tibia and the mean control tibia. The distances are plotted on the reference tibia 3D model. Blue regions indicate that the SMC tibia is smaller at that specific location. Conversely, if the SMC tibia was larger (e.g. caused by a bump or protrusion) the region is colored red. **(Lower part)** Example of an SMC tibia case. Visualisation of the case used to calculate the distances from the reference tibia in the upper part of this image. **(a)** proximal view; **(b)** anterior view; **(c)** posterior view

these distinct morphotypes may help to find the requirements of an optimal therapy strategy for each subgroup of patients. Morphotype is an intrinsic non-modifiable factor in a multifactorial pathogenesis. Especially in high-risk morphotypes modifiable load-related risk factors such as body mass and certain activities could be addressed first in a joint-preserving treatment approach of these pathologies. If the total load reduction proves to be unsatisfactory, local supraphysiological contact stress can be decreased further by unloader braces or a corrective osteotomy in a next stage [22]. Also, in end-stage OA patients with a distinctive morphotype, a morphotype-specific total knee prosthesis might be required. For example, knees with trochlear dysplasia (TD) are often more narrow in the ML direction [14, 15]. Therefore, a narrow variant of a total knee prosthesis may be needed to avoid ML overhang. The current study, which describes the small medial femoral condyle knee as a distinct morphotype, might be related to meniscal pathology. Even when treated according to the gold standards, a meniscal lesion is often associated with the early development of knee OA [23, 24]. Whereas the causal mechanism in the pathogenesis of traumatic tears is clear, degenerative tearing is a multifactorial process. First, genetic predisposition can alter the quality of the meniscal tissue and cartilage [25]. Second, overload due to obesity [26, 27], a high activity level [28] or malalignment [29, 30] (varus or valgus knees) are identified risk factors for early-onset and progression of the degenerative process. In addition to these risk factors, the small medial femoral condyle knee might also be a risk factor leading to overload because of smaller contact areas at the medial compartment. There is evidence that an arthroscopic partial meniscectomy for meniscus lesions might induce adverse biomechanical effects in a yet indefinite subgroup of patients [31, 32]. Based on personal clinical observation of the senior author (PV), the outcome of a medial partial meniscectomy in an SMC knee is characterized by a higher incidence of rapid progressive degeneration, subchondral insufficiency fractures and massive OA in the medial compartment. The decrease in tibiofemoral contact area [32] caused by the removal of meniscus tissue in an SMC knee is most likely associated with a higher increase of contact stress than it would be in a control group knee. This elevated contact stress is then a driving factor for further degeneration of the affected compartment [21, 33, 34]. Further biomechanical and clinical observational studies are necessary to elucidate this mechanism and confirm this hypothesis. Although the knee morphotype may play an important role in the multifactorial degenerative process, it does not explain the complete etiopathogenesis of knee degeneration. In this study two complementary methodologies were applied to evaluate a set of predefined parameters and to evaluate the global shape of the knee models. Where the landmark-based analysis is successful in quantifying several clinically relevant parameters and is easy to interpret, the global shape analysis may detect shape differences which were not captured by the first method. Additionally, it allows for a high level of automation, which makes it a suitable approach to analyse a large number of shapes. For the femur, the results from the global shape analysis were confirmed by the already validated landmark-based 3D analysis. The results of the tibia analysis showed the same trends . However, in contrast with the landmark analysis, the global shape analysis failed to find significantly different regions. This can be explained by the following mechanism, inherent to the two methodologies. The landmark-based analysis looks for differences in certain pre-defined dimensions of the bone. For example: the ML width of the distal femur is the distance between the medial and lateral femoral epicondyle. If the medial epicondyle in the SMC group tends to be located more medially and the lateral epicondyle more laterally, this results in an additive effect for the ML width of the distal

femur in the landmark-based analysis, but not so in the global shape analysis. The main limitations of this study are the small sample size in both groups and the obvious case selection bias, which are intrinsically linked to the very nature of an exploratory study. Most studies of this exploratory design have a limited number of patients to avoid spending too much time in case of a null effect, as the segmentation and analysis process is time-consuming. As a consequence of the small sample size, the control group might not be completely representative for a normal knee anatomy on population level. As already shown for the coronal alignment phenotype concept, there can be large variations on population level, even in healthy knees [35, 10]. It is highly unlikely that all physiologic shape variance of the knee joint was captured in this small control group. Therefore, future research should not only focus on how the pathology deviates from the normal, but also in acquiring larger datasets that can more accurately characterize normal knee morphology. Furthermore, the MRI datasets didn't capture the full lower limb. The addition of the coronal alignment phenotype concept (variation in applied load direction) to the knee morphotype concept (variation in shape and hence also surface area) would be highly interesting to estimate the contact stress (by definition: perpendicular projection of the force divided by the surface area) distribution in both compartments. A clear definition of 'the small medial femoral condyle knee' is not yet fully established, as for the moment there is no evidence of a threshold in terms of a knee shape at which the resulting joint contact stress is not tolerated well anymore. However, this study demonstrated that the SMC knee can be identified on a clinical MRI. The main value of this article for daily clinical practice lies in creating awareness for this morphotype as a risk and prognostic factor in medial compartment degeneration. Early identification of knees at risk might help to start conservative treatment at an earlier stage by addressing modifiable risk factors such as body mass and/or certain activities.

12.5 Conclusion

A new morphotype of the knee demonstrated medial compartment degeneration and was differentiated from a healthy control group based on the following characteristics: a smaller medial femoral condyle and medial tibial plateau; a wider lateral femoral condyle and a wider distal femur on a smaller tibial plateau. This pilot study suggests a role for the SMC knee morphotype in the multifactorial process of medial compartment degeneration.

Bibliography

[1] Ahmad R, Patel A, Mandalia V, Toms A. Posterior Tibial Slope: Effect on, and Interaction with, Knee Kinematics. JBJS reviews. 2016 4;4:(e3) 1-6.

[2] Biedert RM, Netzer P, Gal I, Sigg A, Tscholl PM. The lateral condyle index: a new index for assessing the length of the lateral articular trochlea as predisposing factor for patellar instability. International Orthopaedics. 2011 9;35:1327-31.

[3] Dejour D, Ferrua P, Ntagiopoulos PG, Radier C, Hulet C, Rémy F, et al. The introduction of a new MRI index to evaluate sagittal patellofemoral engagement. Orthopaedics & Traumatology: Surgery & Research. 2013 12;99:S391-8.

[4] Fernández-Jaén T, López-Alcorocho JM, Rodriguez-Iñigo E, Castellán F, Hernández JC, Guillén-García P. The Importance of the Intercondylar Notch in Anterior Cruciate Ligament Tears. Orthopaedic Journal of Sports Medicine. 2015 8;3:2325967115597882.

[5] Smith HC, Vacek P, Johnson RJ, Slauterbeck JR, Hashemi J, Shultz S, et al. Risk Factors for Anterior Cruciate Ligament Injury. Sports Health: A Multidisciplinary Approach. 2012 1;4:69-78.

[6] Springer B, Bechler U, Waldstein W, Rueckl K, Boettner CS, Boettner F. The influence of femoral and tibial bony anatomy on valgus OA of the knee. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 10:DOI: 10.1007/s00167-019057346.

[7] Haver AV, Roo KD, Beule MD, Labey L, Baets PD, Dejour D, et al. The Effect of Trochlear Dysplasia on Patellofemoral Biomechanics. The American Journal of Sports Medicine. 2015 6;43:1354-61.

[8] Hess S, Moser LB, Amsler F, Behrend H, Hirschmann MT. Highly variable coronal tibial and femoral alignment in osteoarthritic knees: a systematic review. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 5;27:1368-77.

[9] Hirschmann MT, Moser LB, Amsler F, Behrend H, Leclerq V, Hess S. Functional knee phenotypes: a novel classification for phenotyping the coronal lower limb alignment based on the native alignment in young non-osteoarthritic patients. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 5;27:1394-402.

[10] Moser LB, Hess S, Amsler F, Behrend H, Hirschmann MT. Native non-osteoarthritic knees have a highly variable coronal alignment: a systematic review. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 5;27:1359-67.

[11] Dejour H, Walch G, Nove-Josserand L, Guier C. Factors of patellar instability: An anatomic radiographic study. Knee Surgery, Sports Traumatology, Arthroscopy. 1994 3;2:19-26.

[12] Iranpour F, Merican AM, Teo SH, Cobb JP, Amis AA. Femoral articular geometry and patellofemoral stability. The Knee. 2017 6;24:555-63.

[13] Zaffagnini S, Grassi A, Zocco G, Rosa MA, Signorelli C, Muccioli GMM. The patellofemoral joint: from dysplasia to dislocation. EFORT Open Reviews. 2017 5;2:204-14.

[14] Peeters W, Haver AV, den Wyngaert SV, Verdonk P. A landmark-based 3D analysis reveals a narrower tibial plateau and patella in trochlear dysplastic knees. Knee Surgery, Sports Traumatology, Arthroscopy. 2020 7;28:2224-32.

[15] Haver AV, Roo KD, Beule MD, Cauter SV, Audenaert E, Claessens T, et al. Semi-automated landmark-based 3D analysis reveals new morphometric characteristics in the trochlear dysplastic femur. Knee Surgery, Sports Traumatology, Arthroscopy. 2014 11;22:2698-708.

[16] Haver AV, Mahieu P, Claessens T, Li H, Pattyn C, Verdonk P, et al. A statistical shape model of trochlear dysplasia of the knee. The Knee. 2014 3;21:518-23.

[17] Gower JC. Generalized procrustes analysis. Psychometrika. 1975 3;40:33-51.

[18] Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods. 2007 5;39:175-91.

[19] Danckaers F, Huysmans T, Lacko D, Ledda A, Verwulgent S, Dongen SV, et al. Correspondence Preserving Elastic Surface Registration with Shape Model Prior. In: 2014 22nd International Conference on Pattern Recognition. IEEE; 2014. p. 2143-8.

[20] Lage-Castellanos A, Martínez-Montes E, Hernández-Cabrera JA, Galán L. False discovery rate and permutation test: an evaluation in ERP data analysis. Statistics in medicine. 2010 1;29:63-74.

[21] Madry H, Luyten FP, Facchini A. Biological aspects of early osteoarthritis. Knee Surgery, Sports Traumatology, Arthroscopy. 2012 3;20:407-22.

[22] Magnussen RA, Dahm DL, Neyret P. In: Osteotomy for Slope Correction Following Failed ACL Reconstruction. Springer New York; 2014. p. 221-6.

[23] Badlani JT, Borrero C, Golla S, Harner CD, Irrgang JJ. The Effects of Meniscus Injury on the Development of Knee Osteoarthritis. The American Journal of Sports Medicine. 2013 6;41:1238-44.

[24] Englund M, Lohmander LS. Meniscectomy and osteoarthritis: what is the cause and what is the effect? Future Rheumatology. 2006 4;1:207-15.

[25] Brophy RH, Zhang B, Cai L, Wright RW, Sandell LJ, Rai MF. Transcriptome comparison of meniscus from patients with and without osteoarthritis. Osteoarthritis and Cartilage. 2018 3;26:422-32.

[26] Englund M, Guermazi A, Gale D, Hunter DJ, Aliabadi P, Clancy M, et al. Incidental Meniscal Findings on Knee MRI in Middle-Aged and Elderly Persons. New England Journal of Medicine. 2008 9;359:1108-15.

[27] Laberge MA, Baum T, Virayavanich W, Nardo L, Nevitt MC, Lynch J, et al. Obesity increases the prevalence and severity of focal knee abnormalities diagnosed using 3T MRI in middle-aged subjects—data from the Osteoarthritis Initiative. Skeletal Radiology. 2012 6;41:633-41.

[28] Thorlund JB, Juhl CB, Ingelsrud LH, Skou ST. Risk factors, diagnosis and non-surgical treatment for meniscal tears: evidence and recommendations: a statement paper commissioned by the Danish Society of Sports Physical Therapy (DSSF). British Journal of Sports Medicine. 2018 5;52:557-65.

[29] Moisio K, Chang A, Eckstein F, Chmiel JS, Wirth W, Almagor O, et al. Varus–valgus alignment: Reduced risk of subsequent cartilage loss in the less loaded compartment. Arthritis & Rheumatism. 2011 4;63:1002-9.

[30] Sharma L, Song J, Dunlop D, Felson D, Lewis CE, Segal N, et al. Varus and valgus alignment and incident and progressive knee osteoarthritis. Annals of the Rheumatic Diseases. 2010 11;69:1940-5.

[31] Pihl K, Ensor J, Peat G, Englund M, Lohmander S, Jørgensen U, et al. Wild goose chase – no predictable patient subgroups benefit from meniscal surgery: patient-reported outcomes of 641 patients 1 year after surgery. British Journal of Sports Medicine. 2020 1;54:13-22.

[32] Zielinska B, Donahue TLH. 3D Finite Element Model of Meniscectomy: Changes in Joint Contact Behavior. Journal of Biomechanical Engineering. 2006 2;128:115-23.

[33] McNulty AL, Guilak F. Mechanobiology of the meniscus. Journal of Biomechanics. 2015 6;48:1469-78.

[34] Segal NA, Anderson DD, Iyer KS, Baker J, Torner JC, Lynch JA, et al. Baseline articular contact stress levels predict incident symptomatic knee osteoarthritis development in the MOST cohort. Journal of Orthopaedic Research. 2009 12;27:1562-8.

[35] Audenaert EA, Pattyn C, Steenackers G, Roeck JD, Vandermeulen D, Claes P. Statistical Shape Modeling of Skeletal Anatomy for Sex Discrimination: Their Training Size, Sexual Dimorphism, and Asymmetry. Frontiers in Bioengineering and Biotechnology. 2019 11;7:302.

Chapter 13

Meniscus Size Differs Between Patient and Donor Populations for Meniscus Allograft Transplantation

Meniscal allograft transplantation is one of the options to treat symptomatic meniscus deficient knees with good treatment response rates [1], even as a salvage procedure [2]. This chapter describes a study comparing the patient population (dysfunctional meniscus) with the donor population (functional meniscus) in meniscus allograft transplantation. It highlights one of the limitations for a widespread clinical implementation: the matching of donor menisci from a restricted supply to patients (acceptors). More importantly, there is a discrepancy in terms of meniscus size between the meniscus deficient patients (smaller medial menisci) and the general donor population (larger medial menisci). This finding is in line with the conclusion of the previous chapter, where the small medial femoral condyle was linked to early onset medial knee degeneration (including degenerative medial menisci) in knees with a neutral alignment (no varus or valgus).

This work was previously published as: *Tabbaa, S. M., Pace, J. L., Frank, R. M., Grammens, J., Verdonk, P. (2023). Meniscus Size Differs Between Patient and Donor Populations for Meniscus Allograft Transplantation. Arthroscopy, sports medicine, and rehabilitation, 5(3), e569–e576. https://doi.org/10.1016/j.asmr.2023.02.009*

13.1 Introduction

Meniscus allograft transplantation is one of the few treatments available and effective for treating patients with the post-meniscectomy syndrome [3, 4]. Adequate size matching of the donor meniscus allograft to the patient's native meniscus is a critical step that can impact a successful surgery and subsequent patient outcomes [5]. Meniscus-sizing methods that rely on imaging or anthropometric data [6] exist, with the Pollard method via radiograph or magnetic resonance imaging (MRI) being the gold standard [3]. Investigations that correlate anthropometric data, such as height, weight,

and sex to meniscal measurements, also have been conducted and may serve as an alternative lower-cost approach for preoperative meniscus size matching [7, 8]. Despite these methods to accurately match meniscus donors to patients, and while as-yet unpublished, there has been an observed phenomenon among meniscus transplant surgeons and tissue banks that certain meniscus sizes are more difficult to procure. At first glance, one would posit that with a large enough of a sample size (donors and patients) that there would be roughly equal distribution of meniscus sizes available. Thus, if there were difficulties in matching certain meniscus sizes, this suggests that these 2 groups are not equal with regards to meniscus size. Size discrepancies between patient and donor populations could explain the shorter supply and concordant longer wait times for patients with specific meniscus sizes and a surplus of donor menisci at other points along the spectrum. As a result, surgeons may also end up accepting a less-than-ideal meniscus transplant that could compromise patient outcomes [5]. A greater understanding of potential differences between meniscus transplant donor and patient populations is necessary if this issue is to be properly addressed.

The purposes of this study are to determine the extent of variability in meniscus size and anthropometric data between donors (supply) and patients (demand), to evaluate potential factors that may contribute to size discrepancies, and to determine whether the discrepancies lead to longer patient wait times. We hypothesized that variability in meniscus size would exist between donors (supply) and patients (demand) and that these discrepancies lead to longer patient wait times.

13.2 Materials and methods

13.2.1 Data collection and analysis

This study was considered exempt from institutional review board approval due to the deidentified data that were analyzed. Meniscus length and width measurements as well as anthropomorphic data were collected and extracted from a large U.S. tissue bank database for both donor and patient pools from 2016 to 2019. Anthropomorphic data included sex, height, weight, and anatomic side. Body mass index (BMI) was calculated from height and weight data. Samples were included if meniscus size and anthropometric data were recorded. Donor and patient samples were removed from the dataset and analysis if the data were incomplete or incorrectly entered into the system. Donor and patient data were further segmented by medial and lateral meniscus to analyze whether distributions varied between patient and donors for lateral and medial menisci.

With regards to donor menisci and per-company protocol, the donor meniscus size (length and width) was measured directly using calipers in situ at the time of processing before harvesting. Meniscus length was defined as the anterior to posterior distance from the anterior most aspect of the anterior horn to the posterior most aspect of the posterior horn. Meniscus width was defined as the distance between the meniscus root and the medial or lateral most aspect of the body of the meniscus.

Patient meniscus size, height, weight, sex, and anatomic side were extracted and collected from meniscus requests, which are standard aspects of the forms filled out by physician

offices when a meniscus transplant request is made. Patient meniscus measurements were determined using radiographs, MRI, or computed tomography scans and methods established by Pollard et al[9]. In summary, the width was calculated by measuring the distance from the peak of the medial or lateral tibial eminence to the medial or lateral tibial epiphyseal margin for the medial and lateral meniscus, respectively. The length was calculated using the lateral view by measuring 70% and 80% of the sagittal length of the proximal tibia that references the tibial tuberosity anteriorly and the posterior aspect of the lateral tibia plateau posteriorly for the lateral and medial meniscus, respectively.

The time to match a patient to a donor meniscus graft was extracted and calculated from the tissue bank company's database. The time to match was determined by the date of the initial patient request and the date of the first allocation or date a donor was matched to the request.

13.2.2 Distribution of meniscus size and area

To determine whether the meniscus sizes varied between patient and donor groups, distribution plots were generated and analyzed. The meniscus size data for both length and width measurements were categorized and segmented by ±0.2 cm, the industry-allowable tolerance for matching meniscus sizes. The average length, width, and area were measured and compared between donor and patient populations. The meniscus area was estimated using the recorded length and width measurements for patient and donors and using the following equation:

$$\text{Estimated meniscus area } (cm^2) = \pi * \frac{\text{width}}{2} * \frac{\text{length}}{2} \tag{13.1}$$

13.2.3 Body Mass to Meniscus Index (BMMI) and Height Over Meniscus Size Index (HMI)

To determine factors that may influence meniscus size and area discrepancies between donors and patients, anthropometric data including height, weight, BMI, sex, and laterality were compared between patient and donor populations. The BMI was calculated using the recorded weight and height in kilograms and meters, respectively. To understand the relative meniscus size, the ratio of the weight to estimated meniscus area (kg/cm^2), or BMMI, was measured and compared between patient and donor pools. Size discrepancies were further investigated by measuring the height to meniscus size index (cm/cm^2). The ratio of HMI was measured and compared between donor and patient pools.

13.2.4 Time to Matching Evaluation

To determine the effect of patient meniscus size on the time to match a donor meniscus, the average time to match was calculated for each segmented meniscus size range (±0.2 cm). Sizes were segmented by increments of 0.2 cm. The time to match a donor meniscus to male and female patients were measured separately for both medial and

lateral menisci. The time to match for each segmented size range was measured and analyzed to identify if specific size ranges take longer to match. In addition, the effect of laterality on time to match was measured and compared between medial and lateral meniscus groups.

13.2.5 Statistical analysis

An a priori power analysis (power of 0.90 and α of 0.05) was performed to determine the sample size needed to detect a statistically significant difference in meniscus size between donor and patient populations. Using the observed means and standard deviations of pilot data samples, we determined that 40 samples per group was sufficient to distinguish differences in meniscus size between patient and donor populations.

All analyses were performed using JMP Pro 12 (SAS Institute, Cary, NC). Data are presented as mean ± standard deviation with P < .05 considered significant. All collected variables were analyzed using descriptive statistics including means, standard deviations, ranges, and frequencies. Meniscus size ranges were categorized in intervals of 0.2 cm and compared between donor and patient populations using χ^2 tests. The average meniscus size (length, width, area) and relative weight and height to meniscus size was calculated and compared between donor and patient populations using independent 2-sample t-tests. Anthropometric data and laterality were compared between donor and patient populations using Fisher exact tests for categorial variables and 2-sample t-tests for continuous variables.

The effect of patient meniscus size on time to match a donor meniscus was determined using one-way analysis of variance for continuous variables and Tukey post hoc tests. In addition, the effect of patient sex, laterality, and medial versus lateral on time to match was determined using a 2-sample t-test, comparing the mean time to match between various factors (i.e., male vs female and left vs right).

13.3 Results

The database query identified 3,218 donor and 704 patient menisci. The final dataset analyzed after the removal of samples with incomplete or incorrect data included 3,189 donor menisci and 576 patient menisci. Overall differences in sex, laterality, and meniscus type were observed between donor and patient pools (Table 3.1). The frequency of male and female sex varied significantly between donor and patient pools (P < .001). The donor pool consisted of a significantly greater male frequency (72.1%) compared with the patient pool, which consisted of 51.0% male patients. The frequency of left and right meniscus and meniscus type (e.g., medial vs lateral) also significantly differed between donor and patient pools. The donor pool consisted of a greater available meniscus frequency from the left knee and the patient pool showed a higher frequency of demand for a meniscus from the right knee.

Table 13.1: Overall Differences Between Donor and Patient Populations.

Medial Meniscus	Donor	Patient	*P* Value
Male, n(%)	2 296 (72.1%)	294 (51.0%)	<.001*
Female, n(%)	889 (27.9%)	282 (49.0%)	
Left, n(%)	1 677 (52.7%)	271 (47.1%)	.013*
Right, n(%)	1 508 (47.4%)	305 (53.0%)	
Lateral, n(%)	1 775 (55.7%)	285 (49.5%)	.006*
Medial, n(%)	1 410 (44.3%)	291 (50.5%)	

*Defines statistical significant with a *P* value lower than .05.

13.3.1 Medial meniscus (donor versus patient populations)

The distribution of medial meniscus size (Fig 3.1) was significantly different between patient and donor pools (P < .05). The patient population showed a significantly greater frequency/need of smaller medial meniscus width (P < .001) and length (P < .001) measurements compared with the donor population availability. In addition, the average patient meniscus length (Table 2) was significantly smaller (ca. 0.3 cm difference) in both the male (P < .001) and female patient population (P < .001) compared with the donor population, which further supports the distribution profile observed with a greater frequency of smaller patient meniscus sizes requested compared with donor size availability.

Sex-specific data for the medial meniscus between donor and patient populations are presented in Table 3.2. The most consistent and significant discrepancies seen were in meniscus length (P < .001), area (P < . 001), and HMI (P < .001) for both male and female populations. Observed differences varied by sex. The male patient population BMMI was significantly greater compared with the donor BMMI (P < .001). The female population showed no difference in BMI or BMMI between donor and patient pools. The frequency of anatomic side (left vs right) was significantly different between the male patient and donor groups (P = .012) with an increase in frequency of the need for a right meniscus for the patient group and a reduction in right meniscus availability from the donor group. No difference was observed in the female population (P = .435).

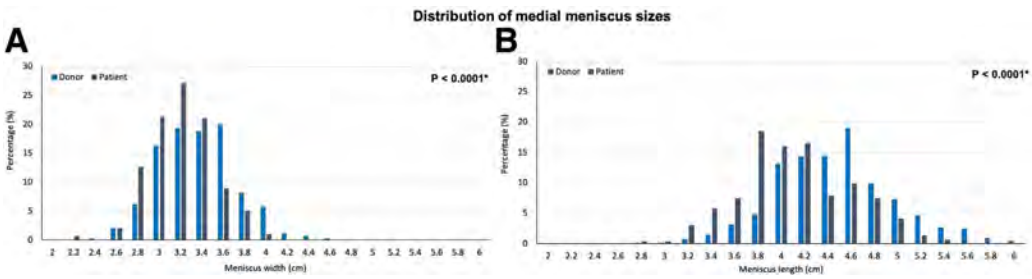


Figure 13.1: Distribution plots showing the mismatch between supply (donors) and demand (patients) for both male and female patients, with the smaller sized medial menisci being in greater demand and undersupplied. Plots show the frequency of donor (blue) and patient (gray) medial meniscus (A) width and (B) length.

Table 13.2: Medial Meniscus Comparison of Meniscus Size, Anthropometric Factors, and Anatomic Side.

Medial meniscus	Donor	Patient	*P* Value
Female medial meniscus			
Meniscus width, cm, mean ± SD	3.04 ± 0.3	2.96 ± 0.2	.002*
Meniscus length, cm, mean ± SD	4.08 ± 0.4	3.83 ± 0.4	<.001*
Meniscus area, cm^2, mean ± SD	9.8 ± 1.8	9.0 ± 1.3	<.001*
Anatomical side, left, n (%)	210 (54.8%)	71 (51.0%)	.435
Anatomical side, right, n (%)	173 (45.2%)	74 (49.0%)	
Height, m, mean ± SD	1.63 ± 0.1	1.66 ± 0.1	<.001*
Weight, kg, mean ± SD	73.5 ± 21.0	68.7 ± 20.5	<.017*
BMI, mean ± SD, kg/cm^2	27.4 ± 7.4	24.8 ± 6.9	<.001*
BMMI, kg/cm^2, mean ± SD	7.6 ± 2.1	7.8 ± 2.4	.474
HMI, cm/cm^2, mean ± SD	17.2 ± 2.8	18.9 ± 2.6	<.001*
Male medial meniscus			
Meniscus width, cm, mean ± SD	3.4 ± 0.3	3.3 ± 0.3	<.001*
Meniscus length, cm, mean ± SD	4.6 ± 0.5	4.3 ± 0.4	<.001*
Meniscus area, cm^2, mean ± SD	12.3 ± 2.2	11.2 ± 1.7	<.001*
Anatomical side, left, n (%)	557 (54.2%)	63 (43.1%)	<.012*
Anatomical side, right, n (%)	479 (45.8%)	83 (56.9%)	
Height, m, mean ± SD	1.77 ± 0.1	1.79 ± 0.1	<.001*
Weight, kg, mean ± SD	83.9 ± 22.9	87.6 ± 19.8	.039*
BMI, mean ± SD, kg/cm^2	26.7 ± 6.7	27.0 ± 5.6	.455
BMMI, kg/cm^2, mean ± SD	6.9 ± 1.9	7.8 ± 1.8	<.001*
HMI, cm/cm^2, mean ± SD	14.8 ± 2.5	16.2 ± 2.6	<.001*

BMI, body mass index; BMMI, body mass to meniscus index; HMI, height to meniscus index; SD, standard deviation.
*Defines statistical significant with a *P* value lower than .05.

13.3.2 Lateral meniscus (donor versus patient populations)

The distribution of lateral meniscus size was significantly different between patient and donor populations for meniscus length (P < . 001). Although discrepancies were observed between donor and patient populations for meniscus width, this measurement was not significant (P = .084). A greater patient demand (Fig 3.2 A and B) of larger lateral meniscus sizes were observed.

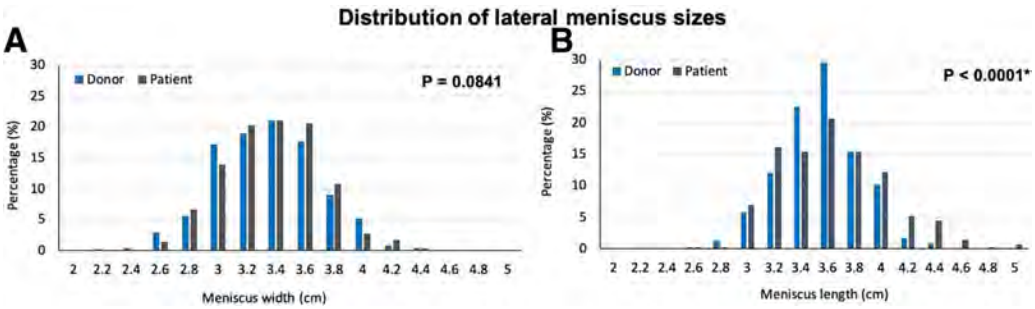


Figure 13.2: Distribution of donor (blue) and patient (gray) lateral meniscus sizes. Lateral meniscus width (A) and length (B).

Sex-specific data for the lateral meniscus between donor and patient populations are presented in Table 3.3. Although meniscus width and length were significantly larger for female patients needing a meniscus (P = .002 , .001), it was not as dramatic as the difference between male donors and patients (P < .001). Male patients were taller than their donor counterparts (P < .001), whereas female patients had a lower BMI (P = .0002) and BMMI (P < .001). The male patients had similar BMI to the donor males but lower BMMI (P = .047). Both male and female patients showed significantly lower HMI compared with the donor group. Consistent with the medial meniscus findings, the frequency of anatomic side was significantly different for the male population (P = .034) and similar for the female population (P = .705).

To further understand the conflicting HMI (greater average HMIs for patients needing a medial meniscus compared with donors vs lower average HMIs for patients needing a lateral meniscus compared to donors), findings between the medial and lateral meniscus, a scatter plot of meniscus area vs patient or donor height was generated (Fig 3.3). The medial meniscus plot (Fig 3.3A) demonstrated patients have a similar height distribution to donors but a smaller medial meniscus size. The lateral meniscus plot depicted similar height and meniscus size distributions for both donor and patient pools (Fig 3.3B).

Table 13.3: Lateral Meniscus Comparison of Meniscus Size, Anthropometric Factors, and Anatomic Side.

Lateral meniscus	Donor	Patient	*P* Value
Female lateral meniscus			
Meniscus width, cm, mean ± SD	3.0 ± 0.3	3.1 ± 0.3	.002*
Meniscus length, cm, mean ± SD	3.3 ± 0.3	3.4 ± 0.3	.001*
Meniscus area, cm^2, mean ± SD	7.8 ± 1.1	8.2 ± 1.3	<.001*
Anatomical side, left, n (%)	253 (50.0%)	71 (51.8%)	.705
Anatomical side, right, n (%)	253 (50.0%)	66 (48.2%)	
Height, m, mean ± SD	1.6 ± 0.1	1.7 ± 0.1	.121
Weight, kg, mean ± SD	73.9 ± 22.4	67.9 ± 20.3	.003*
BMI, mean ± SD, kg/cm^2	27.4 ± 7.8	24.8 ± 6.9	<.001*
BMMI, kg/cm^2, mean ± SD	9.6 ± 3.0	8.4 ± 2.5	<.001*
HMI, cm/cm^2, mean ± SD	21.5 ± 3.0	20.5 ± 2.8	<.001*
Male lateral meniscus			
Meniscus width, cm, mean ± SD	3.4 ± 0.3	3.5 ± 0.3	<.001*
Meniscus length, cm, mean ± SD	3.6 ± 0.3	3.8 ± 0.4	<.001*
Meniscus area, cm^2, mean ± SD	9.7 ± 1.5	10.5 ± 1.6	<.001*
Anatomical side, left, n (%)	657 (51.8%)	63 (42.6%)	.034*
Anatomical side, right, n (%)	479 (45.8%)	83 (56.9%)	
Height, m, mean ± SD	1.77 ± 0.1	1.80 ± 0.1	<.001*
Weight, kg, mean ± SD	83.3 ± 22.0	86.6 ± 25.2	.133
BMI, mean ± SD, kg/cm^2	26.5 ± 6.4	26.6 ± 7.2	.865
BMMI, kg/cm^2, mean ± SD	8.7 ± 2.3	8.3 ± 2.4	.047*
HMI, cm/cm^2, mean ± SD	18.7 ± 2.5	17.4 ± 2.2	<.001*

BMI, body mass index; BMMI, body mass to meniscus index; HMI, height to meniscus index; SD, standard deviation.
*Defines statistical significant with a *P* value lower than .05.

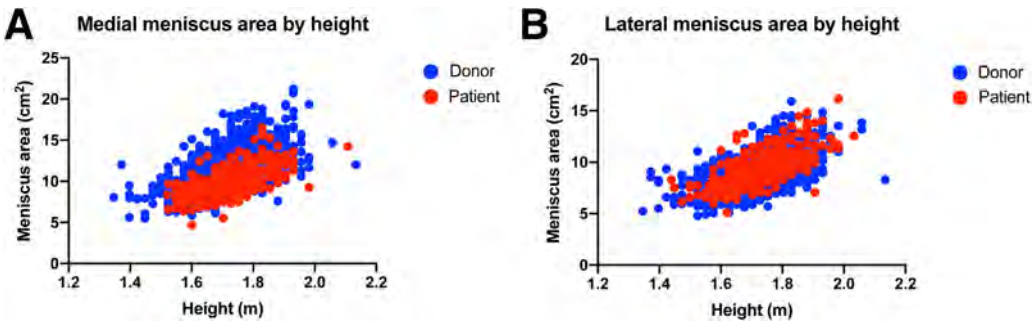


Figure 13.3: Scatter plot of meniscus area vs patient or donor height for (A) medial and (B) lateral meniscus.

13.3.3 Time to Match a Patient Request to Donor Graft

The effect of patient sex, and anatomic side on time to match a donor graft was analyzed for medial and lateral meniscus (Table 3.4). The anatomical site influenced the time to match for both medial and lateral meniscus. Sex had no effect on time to match.

The average time to match a patient lateral meniscus was influenced most greatly by the patient length measurement for both medial ($P < .05$) and lateral ($P < .05$) meniscus (Fig 3.4). The average time to match a lateral meniscus was significantly increased for larger meniscus lengths (>4.4 cm) (Fig 3.4B). The medial meniscus analysis showed a different effect where the smaller meniscus length (<3.0 cm) increased the time needed to match a donor graft (Fig 3.4A).

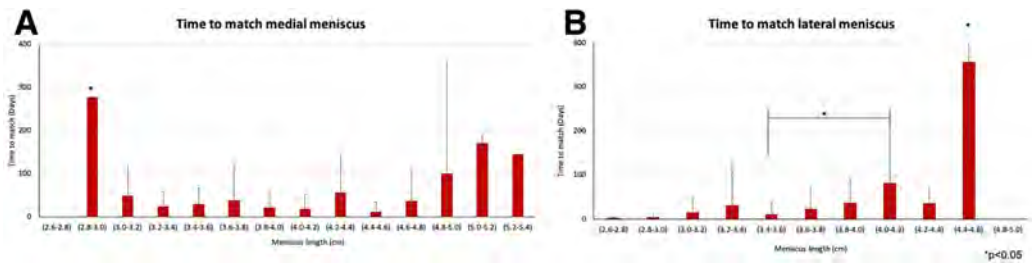


Figure 13.4: The effect of patient meniscus length on time to match a donor graft. (A) Medial meniscus time to match for various meniscus lengths. (B) Lateral meniscus time to match for various meniscus lengths. $P < .05$.

13.4 Discussion

The most important finding of this study is confirmation of the mismatch between donor and patient meniscus sizes. Outcomes from the lateral meniscus analysis showed a high patient demand for larger lateral meniscus sizes, which were not met by the donor population. Interestingly, the mismatch was converse for medial meniscus where

Table 13.4: Effect of Sex and Anatomical Site on the Average Time to Match a Donor Graft.

Category	Average Time to Match	*P* Value
Medial meniscus		
Sex, male, d, mean ± SD	40.0 ± 96.7	.230
Sex. female, d, mean ± SD	28.1 ± 54.9	
Anatomical site, left, d, mean ± SD	25.1 ± 56.8	.073
Anatomical site, right, d, mean ± SD	42.4 ± 94.0	
Lateral meniscus		
Sex, male, d, mean ± SD	35.9 ± 82.3	.208
Sex. female, d, mean ± SD	23.3 ± 71.2	
Anatomical site, left, d, mean ± SD	17.1 ± 55.4	.013*
Anatomical site, right, d, mean ± SD	41.2 ± 91.4	

SD, standard deviation
*Defines statistical significant with a *P* value lower than .05.

greater patient demand for smaller meniscus sizes and a lower donor supply was identified. These data were further corroborated by the time to match where larger lateral meniscus and smaller medial meniscus sizes increased the patient wait time to identify a donor match. Further, there is significant sex differences between donors and patients and in laterality needs between sexes. Male patients compromised a dominant majority of the donor menisci pool (72.1%) while being only 51% of the population of patients who request a meniscus transplant. Male patients requested a significantly greater percentage of meniscus grafts for the right knee while there was much more availability of meniscus grafts from the left knee.

Adequate size matching of patient and donor meniscus before meniscal allograft transplantation (MAT) is an important factor that can influence biomechanics and patient outcomes [5, 10, 11]. Various sizing methods spanning models involving demographic data to techniques using radiographic and MRI have been developed to avoid size discrepancies and improve the matching of donor grafts to patients [8, 12]. Although a number of studies have evaluated mismatch attributed to various sizing methods, there are no previous studies investigating how well the source donor meniscus tissue meets the demands of the patients with meniscal deficiencies. The mismatch in distribution between donors and patient meniscus sizes can impact the availability of grafts for patient care. This study identified a lateral meniscus mismatch, which could be attributed to a larger male patient subpopulation with meniscal deficiency who requires a large-sized meniscus graft. This larger male patient demographic is outside the normal distribution of typical donors leading to a lower frequency of grafts available to serve this patient demographic. Similarly, a high frequency of small medial meniscus grafts is in demand to meet the clinical needs for a smaller patient population, which is also outside the normal distribution of medial meniscus donors. This work identifies a clinical need for specific patient populations

who have limited opportunity to obtain a donor match.

Anthropometric data (sex, laterality, height, weight) was analyzed to understand the factors that may contribute to the meniscus size discrepancy identified between donors and patients. Although this study did not investigate age differences between donor and patient pools, it is unlikely that age would introduce any bias into this study as prior literature reports that most patients who undergo MAT are between 18 and 50 years [13, 14, 15, 16]. Overall differences in sex, laterality, and meniscus type were observed between donor and patient pools. The significantly greater male percentage identified in the donor pool is likely attributed to a large percentage of male donors that are more frequently trauma victims. Further investigation into the exact cause of death in our donor pool could confirm this assertation. Studies investigating organ and tissue donor characteristics have classified a number of deceased donors as trauma donors, or donors with cause of death that was not designated as natural causes [17, 18]. Ackerman et al [17]. report the characteristics of trauma donors from 2007 to 2016 [13]. The majority of trauma donors identified in this study were male (74.3%) with a mean age of 31.1 years. The trauma donors comprised a younger and healthier population of donors compared with the nontrauma donor counterparts [17]. In addition, the tissue bank providing the meniscus allograft data for this study defines specific donor criteria that influence the donors that will be included or excluded for meniscus allograft tissue. The tissue bank specifies donors between the ages of 12 and 45 years with healthy and intact menisci. Because of the age and meniscus tissue health requirements, trauma donors are likely the main source for meniscus allograft tissue. This was confirmed by the significantly greater proportion of male donors, 72.1%, identified by this study. The greater male proportion of donors may contribute to the mismatch observed for the medial meniscus, where the demand for smaller meniscus sizes outstripped the supply from the donor pool and in which the patient pool was nearly half female. The lower frequency of small meniscus sizes may limit the availability of meniscus donor grafts for female patients or skeletally immature patients [19].

The mismatch in frequency of smaller-sized medial menisci between donor and patients was observed for both male and female populations. Sex-specific anthropometric data analysis was conducted to further understand the size discrepancy. Interestingly, the frequency of anatomic side played a role with the male population, but the not the female population. There was an increased demand for right meniscus grafts for the male patient group and a reduction in availability from the donors. This suggests leg dominance influenced both the demand and supply of the grafts. Male donors and patients both favor the right side creating a mismatch in availability. Another important factor potentially influencing the medial meniscus mismatch is the HMI. The HMI were significantly greater for both male and female patients compared with the donor population. This suggests that patients and donors have similar height distributions, but the patients have a smaller meniscus size leading to significantly greater HMI. Based on these findings, we hypothesized that individuals with a smaller medial compartment relative to their overall height or BMI are more susceptible to being symptomatic in the face of medial meniscus deficiency. This hypothesis was further supported by the scatter plot of meniscus area versus patient or donor height, which demonstrates patients both have a similar height distribution to donors but a smaller medial meniscus size.

The lateral meniscus size distribution findings differed from the medial meniscus. The

mismatch between donors and patients was observed for larger meniscus sizes. The clinical significance of this mismatch is the limited availability of donor lateral meniscus allografts to meet the demands of larger patients. Anthropometric data were analyzed to understand factors that may contribute to this size discrepancy. Similar to the medial meniscus findings, the leg dominance factor was observed for only male patients. In addition, donor height played an important role in the mismatch observed. Patients were on average significantly taller than the donor pools. This is consistent with patient demographics reported for MAT. The HMI hypothesis, however, was not observed for the lateral meniscus. We believe the primary contributor to the lateral meniscus discrepancy is the difference in height distributions between patients and donors. Patients are skewed to taller heights, where donors have a normal distribution.

As evidenced in this study, the observed mismatches between patient and donor pools led to delays in treatment or longer wait times to identify a donor match. Larger-sized lateral meniscus and smaller-sized medial meniscus graft requests increased the time to identify a donor match. Although there is no immediate clinical action that can be taken to remedy this situation, this does provide treating physicians with information that can be used when discussing MAT with patients. Although most patients will be able to find a donor match in a reasonable amount of time, specific subsets of patients are at risk of extended wait times for a graft. However, while this work highlights shortages of certain meniscus sizes, there is a converse excess supply of other meniscus graft sizes, specifically smaller lateral meniscus and larger medial meniscus grafts. Given these observations, one strategy to overcome this limitation is to consider the use of a medial meniscus allograft for a lateral meniscus application or vice versa. To the extent of our knowledge, there are no published studies investigating the feasibility of using a donor medial allograft for a lateral meniscus recipient. Laboratory, animal, and clinical studies would be needed to justify the use of medial-to-lateral meniscus or lateral-to-medial-meniscus transplantation. Another option would be to consider segmental meniscus transplantation when the meniscus deficiency is not global. Early animal work in this area has been mixed repair outcomes but highlights that such a need exists[20].

We acknowledge several limitations of this study. First, this study did not analyze the age of donor and patient pools due to restrictions obtaining this information consistently from a database which could introduce potential bias. Another limitation involved the methods for meniscus size measurement. The patient meniscal measurements were determined by outside sources that used various imaging modalities including radiographs, MRI, or computed tomography scans whereas donor meniscus measurements were made directly with hand calipers. The various types of scans and lack of standardization for how measurements were taken for patients may lead to variability within the study. Lastly, it is possible that interobserver error may affect the donor meniscus measurements recorded using calipers in situ at the time of processing.

13.5 Conclusion

This analysis demonstrates variations in frequency of meniscus sizes between donor and patient populations. This variation is attributed to differences in anthropometric

data between patient and donor populations. This work identifies a mismatch between demand and supply for certain patient sizes contributing to longer times to match.

Bibliography

[1] Wagner KR, Kaiser JT, Hevesi M, Cotter EJ, Gilat R, Meeker ZD, et al. Minimum 10-Year Clinical Outcomes and Survivorship of Meniscal Allograft Transplantation With Fresh-Frozen Allografts Using the Bridge-in-Slot Technique. AMERICAN JOURNAL OF SPORTS MEDICINE. 2023 SEP;51(11):2954-63.

[2] Romandini I, Grassi A, Andrea Lucidi G, Filardo G, Zaffagnini S. 10-Year Survival and Clinical Improvement of Meniscal Allograft Transplantation in Early to Moderate Knee Osteoarthritis. AMERICAN JOURNAL OF SPORTS MEDICINE. 2024 JUL;52(8):1997-2007.

[3] Samitier G, Alentorn-Geli E, Taylor DC, et al. Meniscal allograft transplantation. Part 1: Systematic review of graft biology, graft shrinkage, graft extrusion, graft sizing, and graft fixation. Knee Surgery. Sport Traumatol Arthrosc. 2014;23:310-22.

[4] Berhouet J, Marty F, Rosset P, Favard L. Meniscus matching: Evaluation of direct anatomical, indirect radiographic, and photographic methods in 10 cadaver knees. Orthop Traumatol Surg Res. 2013;99:291-7.

[5] Dienst M, Greis PE, Ellis BJ, Bachus KN, Burks RT. Effect of lateral meniscal allograft sizing on contact mechanics of the lateral tibial plateau: An experimental study in human cadaveric knee joints. Am J Sports Med. 2007;35:34-42.

[6] Shaffer B, Kennedy S, Klimkiewicz J, Yao L. Preoperative sizing of meniscal allografts in meniscus transplantation. Am J Sports Med. 2000;28:524-33.

[7] Netto A dos S, Kaleka CC, Toma MK, et al. Should the meniscal height be considered for preoperative sizing in meniscal transplantation? Knee Surg Sport Traumatol Arthrosc. Knee Surg Sport Traumatol Arthrosc. 2018;26:772-80.

[8] Van Thiel GS, Verma N, Yanke A, Basu S, Farr J, Cole B. Meniscal allograft size can be predicted by height, weight, and gender. Arthros copy. 2009;25:722-7.

[9] Pollard ME, Kang Q, Berg EE. Radiographic sizing for meniscal transplantation. Arthroscopy. 1995;11:684-7.

[10] Kaleka CC, Netto AS, Silva JC, et al. Which are the most reliable methods of predicting the meniscal size for transplantation? Am J Sports Med. 2016;44:2876-83.

[11] Stevenson C, Mahmoud A, Tudor F, Myers P. Meniscal allograft transplantation: Undersizing grafts can lead to increased rates of clinical and mechanical failure. Knee Surg Sport Traumatol Arthrosc. 2019;27:1900-7.

[12] Yoon JR, Kim TS, Wang JH, Yun HH, Lim H, Yang JH. Importance of independent measurement of width and length of lateral meniscus during preoperative sizing for meniscal allograft transplantation. Am J Sports Med. 2011;39:1541-7.

[13] Cole BJ, Dennis MG, Lee SJ, et al. Prospective evaluation of allograft meniscus transplantation: A minimum 2-year follow-up. Am J Sports Med. 2006;34:919-27.

[14] Verdonk PCM, Demurie A, Almqvist KF, Veys EM, Verbruggen G, Verdonk R. Transplantation of viable meniscal allograft. J Bone Joint Surg Am. 2005;87:715-24.

[15] Vundelinckx B, Bellemans J, Vanlauwe J. Arthroscopically assisted meniscal allograft transplantation in the knee: A medium-term subjective, clinical, and radiographical outcome evaluation. Am J Sports Med. 2010;38:2240-7.

[16] Kim C, Il BS, Kim JM, et al. Medial and lateral meniscus allograft transplantation showed no difference with respect to graft survivorship and clinical outcomes: A comparative analysis with a minimum 2-year follow-up. Arthroscopy. 2020;36:3061-8.

[17] Ackerman A, Clark D, Lipinska J, Chung B, Whiting J. Organ donation after trauma: A 30-year review. J Trauma Acute Care Surg. 2019;87:130-3.

[18] Nygaard CE, Townsend RN, Diamond DL. Organ donor management and organ outcome: A 6-year review from a Level I trauma center. J Trauma. 1990;30:728-32.

[19] Riboh JC, Tilton AK, Cvetanovich GL, Campbell KA, Cole BJ. Meniscal allograft transplantation in the adolescent population. Arthroscopy. 2016;32:1133-40.

[20] Strauss E, Caborn DNM, Nyland J, Horng S, Chagnon M, Wilke D. Tissue healing following segmental meniscal allograft transplantation: a pilot study. Knee Surg Sport Traumatol Arthrosc. 2019;27:1931-8.

Chapter 14

3D bone morphology is a risk factor for medial post-meniscectomy syndrome: a retrospective cohort study

This chapter describes the comparative study of the knee bone morphology between a cohort of successful partial meniscectomy patients with a cohort of medial post meniscectomy patients. Furthermore, some initial steps towards a morphotype-based predictive model are explored.

This work was previously published as: *Grammens, J., Van Haver, A., Danckaers, F., Vuylsteke, K., Sijbers, J., Mahluf, L., Angele, P., Kon, E., Verdonk, P., MEFISTO WP1 Group (2024). Three-dimensional bone morphology is a risk factor for medial postmeniscectomy syndrome: A retrospective cohort study. Journal of Experimental Orthopaedics, 11(3), e12090. https://doi.org/10.1002/jeo2.12090*

14.1 Introduction

Amongst possible surgical treatment options for medial meniscal tears, arthroscopic partial medial meniscectomy (APMM) is one of the most regularly performed knee surgery worldwide [1, 2, 3, 4, 5]. Recent insights have resulted in an evolution of its indications [6] and there is a paradigm shift towards preserving the meniscus to the greatest extent possible [7]. Current guidelines [8, 9] do not recommend APMM as a first-line treatment, but prefer meniscal repair or conservative treatment instead. It might however serve as an alternative when the latter two treatment options are not applicable (complex tears, tears with high degree of degeneration, flap tears or nonreducible bucket handle tears) or when response to meniscal repair or conservative treatment has been unsatisfactory. A significant subset (6-25%) of partial meniscectomy patients experience persistent or recurrent pain within 1 to 2 years after APMM [10, 11], also known as the post-meniscectomy syndrome. These patients typically suffer from a dull and nagging pain in the operated knee compartment, often accompanied by transient joint effusions [12]. Several studies have established common risk factors for inferior APMM outcome [13], including age, obesity, cartilage status, coronal

malalignment [14], and the extent of the meniscectomy [15]. These risk factors are all related to a mismatch between the applied load (obesity, activity level, coronal malalignment) and a reduced resilience to resist and endure that load (meniscus dysfunction or cartilage loss). An often overlooked but potentially significant risk factor for medial post-meniscectomy syndrome (MPMS) is 3D knee morphology. Previous publications on medial knee compartment morphology indicated a potential link between a small medial femoral condyle and early degeneration of the medial meniscus [16, 17]. The relationship between specific bony knee morphology variations and certain pathologies such as cruciate ligament lesions [18] and patellar instability [19] has already been demonstrated. Although bony morphology is non-modifiable, it is of great importance to detect and acknowledge this factor, as it may have an influence on surgery outcome [20]. A commonly used method to analyze bone morphology is measuring several anatomical landmark-based distances and angles, either on the 3D bone models or directly on the medical images. While very straightforward and easy to visualize, this approach suffers from some limitations. Firstly, it only captures information at discrete, pre-defined anatomical landmarks, thereby neglecting the complexity of shape variation patterns over the entire bone surface. Secondly, the choice of morphological parameters or measurements is subjective and might be prone to a selection bias, potentially overlooking essential morphological features. Statistical shape modeling offers a powerful alternative that overcomes these limitations. By analyzing large datasets of 3D bone models (e.g. from MRI scans), statistical shape modeling involves the creation of a smart shape atlas, that captures the average bone shape and its main modes of shape variation in a data-driven way. It eliminates the need for subjectively chosen landmarks and instead analyzes the entire articular bone surface at once. As a result, it will lead to a deeper understanding of the complex interplay between distinct morphological features in the context of APMM outcome. Statistical shape modeling has been around for several years [21], but only recently found its way to the field of orthopedics [22, 23, 24, 25]. In the present study, it allows to quantitatively compare the femoral and tibial bone shapes between two groups of meniscectomy patients and to extract potential morphological predictors for clinical response to APMM. The purpose of this study was to investigate if adult APMM patients who develop pain symptoms (MPMS), demonstrate different bony knee morphology, compared to APMM patients who don't develop pain symptoms, within a follow up of 2 years after APMM. Based on our previous findings [16, 17], related to a small medial femoral condyle morphotype as a risk factor in the multifactorial process of medial compartment knee pain, we hypothesize the MPMS knees to have a smaller medial femoral condyle. Finally, this study aimed to evaluate a predictive model for APMM outcome with bony morphology as predictor variables.

14.2 Materials and methods

This study is a multicenter, retrospective case-control study. Three high-volume orthopedic centra specialized in knee pathology from Antwerp, Milan and Regensburg participated in this study. IRB approval was obtained from all local ethical committees (Comité voor medische ethiek AZ Monica and UZA: study B300201941743, Ethikkommission an der Universität Regensburg: reference 19-1621-101, IRCCS Instituto Clinico Humanitas: study authorization n. 2515) and informed consent was

obtained from all patients prior to their inclusion.

14.2.1 Patient selection and study design

Patients were eligible for the study if they were between 18 and 70 years old, had a primary medial meniscus lesion, for which meniscal repair or conservative treatment was not applicable and hence an APMM was indicated and performed by an expert surgeon (N=844). Predefined exclusion criteria were unavailable pre-operative MRI, inability to communicate, an unstable knee (IKDC grade C or D), patellar instability or trochlear dysplasia, limited knee range of motion (IKDC grade C or D), cartilage lesions (grade IV and larger than 2cm, non-focal), coronal malalignment (as judged clinically), concomitant discoid meniscus, morbid obesity (BMI > 35), a history of meniscus repair or major lower limb surgery prior to the meniscectomy, septic or rheumatoid arthritis, neurological disorders, posterior cruciate ligament repair or reconstruction, insufficiency fractures or avascular necrosis, plica syndrome or less than 2mm intact medial meniscal rim left intraoperatively. Finally, 443 patients were eligible after screening of their hospital records, of which 42 had an MRI of insufficient quality (movement artefacts or slice thickness > 4mm) and 161 did not consent study participation or did not complete the KOOS questionnaire. As pain is the primary symptom for diagnosis of the post-meniscectomy syndrome, the KOOS pain subscore [26, 27] was used to split the APMM patients in two groups: a first group who showed a good clinical response to APMM (further referred to as R group) and a second group who developed medial post-meniscectomy syndrome (further referred to as MPMS group). A power analysis defined the required sample size as 120 patients per group. The KOOS pain score threshold for stratification of the subgroups was set at 75, based on the patient acceptable symptom state, as calculated by Agarwalla et al [28]. The R group included 120 patients with a KOOS pain score > 75 and the MPMS group included 120 patients with a failed clinical outcome, defined as KOOS pain score < 75. The predefined total study sample size (N=240) was considered adequate to build a robust SSM covering population variance [29]. The patient selection procedure is summarized in the CONSORT diagram in Figure 12.1. All patients were first approached by phone. Documents (informed consent and KOOS questionnaire) were sent following oral consent to participate into the study. Upon attaining the predefined sample size in one group (n=120) no additional patients were included in that group and attempts to contact patients with incomplete KOOS questionnaire or informed consent were stopped when both groups were complete (n=240).

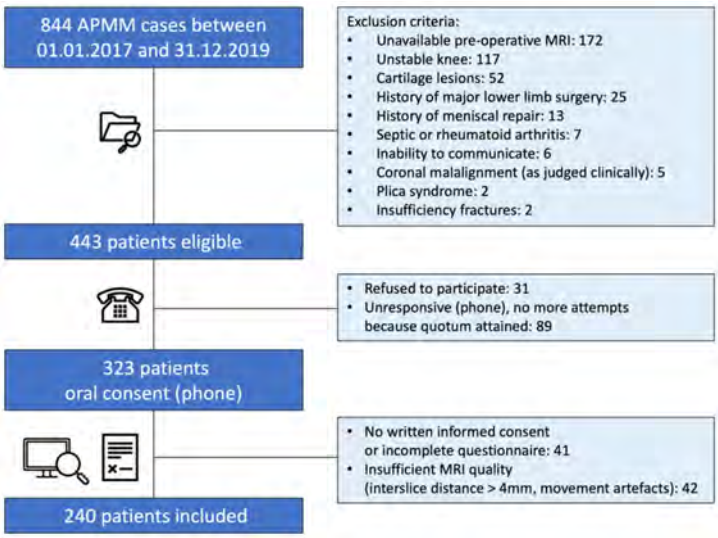


Figure 14.1: Consolidated Standards of Reporting Trials flow diagram for subject enrolment procedure. All subjects underwent an arthroscopic partial medial meniscectomy (APMM). MRI, magnetic resonance imaging.

14.2.2 Data collection

Imaging, demographic and clinical data were collected. The imaging data consisted of the pre-operative MRI scans used for diagnosis of the meniscal lesion and was extracted from the hospital PACS. MRI scans were evaluated by one experienced researcher for motion artefacts and slice thickness < 4 mm. A typical pre-operative MRI (1.5T or 3T) protocol included the following sequences: coronal T1-weighted (3.5 mm slice thickness, 0.7x0.7mm in-plane resolution), coronal proton density weighted (PD) (3.5 mm slice thickness, 0.5x0.5mm in-plane resolution), sagittal PD (3.5 mm slice thickness, 0.5x0.5mm in-plane resolution) and axial T2- or intermediate-weighted (3 mm slice thickness, 0.5x0.5mm in-plane resolution) images. as e and clinical data included patient sex, age, weight and height. BMI was calculated from patient weight and height as follows: patient weight (kg) divided by the square patient height (m). Cartilage status (modified Outerbridge classification [30]) was extracted from the surgery reports and verified on the preoperative MRI scans. The KOOS questionnaire [27] at 2 years of follow-up served as patient-reported outcome measure to evaluate response to treatment (RTT).

14.2.3 Generation of patient-specific 3D bone models

Pre-operative MRI scans were extracted from the PACS system in DICOM format and loaded into Mimics 23.0 (Materialise NV, Leuven, Belgium). Following the standard knee scan protocols, sequences in the three perpendicular anatomical planes were available. Using at least two MRI sequences with a perpendicular acquisition plane, distal femur and proximal tibia were segmented into two separate 3D models, each of them consisting

of bone and cartilage united. The projected contours of the resulting 3D models were doublechecked on all available sequences and finetuned manually using the "Contour edit" tool of the software package. 3D models of the distal femur and proximal tibia were then saved as triangular meshes and further used to perform the morphological analysis.

14.2.4 Statistical shape model: data-driven morphology description

In total, three statistical shape models (SSMs) were built from the 3D bone models: one separate SSM for the isolated distal femur, one SSM for the isolated proximal tibia [31], and a combined SSM of the tibiofemoral joint in full extension following the methodology described by Audenaert et al [32]. The first step in building an SSM is the establishment of anatomical correspondences between the patient-specific bone shapes. This involves the automatic identification and matching of the same anatomical landmarks and regions across the different bone shapes in the dataset. Practically, this was achieved through an iterative process of rigid and elastic deformations as described and validated by Danckaers et al. [31]. Briefly, a template triangular bone mesh of the distal femur or proximal tibia was gradually deformed to the bone shape of the patients, resulting in a mesh with a dense set of pseudo-landmarks that share a consistent anatomical meaning across all patients. Next, all patient bone shapes were rigidly aligned (rotations, translations) to the same position by minimizing the sum of distances between all corresponding points. To reduce any variability induced by the MRI scan field of view, only the distal portion of the femur and proximal portion of the tibia was selected for further analysis (blue region in Figure 12.2).

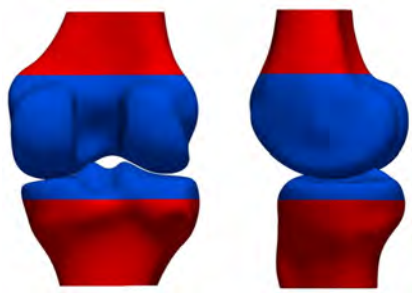


Figure 14.2: Definition of the distal femur and proximal tibia to be included in the shape analysis (blue region). The red surface region was neglected in the shape analysis. Left: posterior view. Right: medial view.

The previously established anatomical correspondences ensure that the same anatomical region is selected for all patients. A custom Python script was written to construct the SSM. The mean bone shape was calculated and principal component analysis (PCA) resulted in the main modes of shape variation [21]. PCA is a commonly used mathematical method for dimensionality reduction, which transforms the original variables (numerous 3D model point coordinates x,y and z for each bone shape) linearly into a set of new variables (confined number of modes of shape variation), ordered by decreasing magnitude (explained shape variance). Together, the mean bone shape and

the modes of shape variation define the SSM and they can describe any shape of the same nature as a weighted sum of those modes. The weight factors for each mode of shape variation are then called principal component (PC) weights. Given the SSM, the PC weights for each mode of shape variation are then used to characterize the patient-specific bone morphology. The SSM construction is summarized and illustrated in Fig. 12.3.

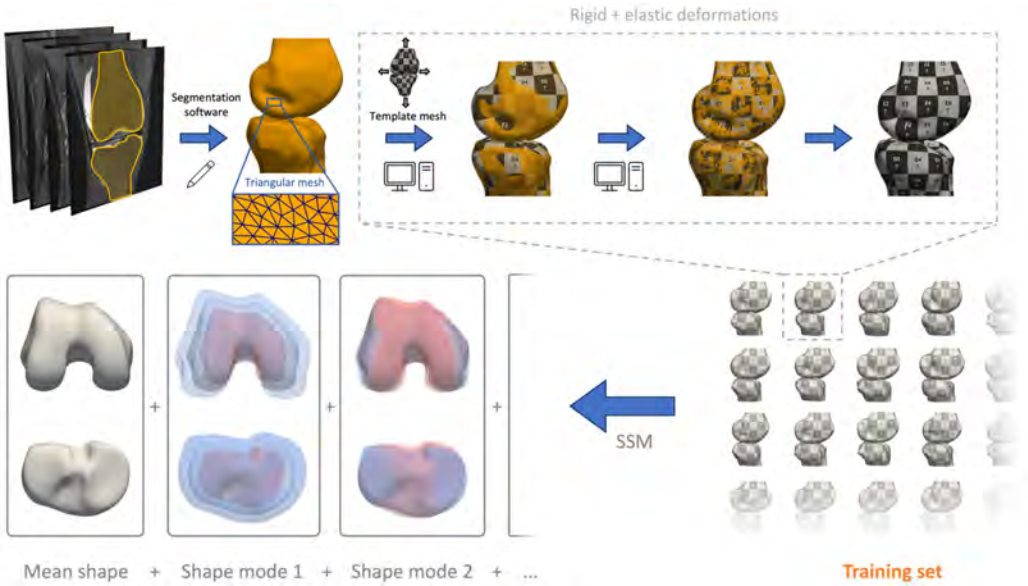


Figure 14.3: Starting from the medical images, three-dimensional models were manually created in the segmentation software. Next, anatomical correspondences are computed by rigidly and elastically deforming a template mesh (visualised with checker pattern). These deformed template meshes (=training set) are then further used to construct the statistical shape model (SSM).

In the combined SSM, the adopted method of Audenaert et al. [32] realigned the individual bones to an average (neutral) relative position before inclusion in the SSM, thereby removing any potential relative positional information (e.g. induced by patient positioning in the scanner). Without this realignment, any positional variation (e.g. flexion/extension, varus/valgus, internal/external rotation) would also be captured in the main modes of shape variation, thereby resulting in a less compact SSM. For all three SSM's in this study, the modes of shape variation were defined on the complete dataset of 240 knees. Model compactness and generalization were evaluated as performance metrics for all three SSM's [29].

Model compactness

Compact shape models can describe any new shape instance with as little modes of shape variation as possible [33]. SSM compactness is described as the cumulative explained variance in function of the number of modes. The higher this value, the fewer modes are

needed for the SSM to describe shape variation at population level and increase model performance. A cumulative explained variance ratio of more than 98% was aimed for when choosing the appropriate number of modes.

Model generalizability

For matters of future clinical applications, the SSM should generalize to unseen shapes of the same nature [33]. That is, it should be able to describe unseen bone shapes to a certain level of accuracy. Repeated leave-one-out cross-validation experiments were performed in order to assess this capability for an increasing number of modes of shape variation. The generalization metric was then defined as the average description error (RMSE) over all experiments.

SSM in relation to clinical data and response to treatment

PC weights for the specific modes of shape variation were calculated for all patients in both R and MPMS groups. The PC weights for the first three modes of shape variation were then compared between the R and MPMS group and a physical meaning was assigned to those modes (results section B. Knee morphology comparison between R and MPMS group). In a next step, correlations between demographic, clinical and morphology variables (PC weights) were assessed (section C. Correlation analysis between demographic, clinical and morphology variables). Finally, a predictive algorithm for RTT based on knee morphology was trained and evaluated (section D. Prediction of response to treatment based on knee morphology).

14.2.5 Statistical analysis

All statistical analyses were performed in RStudio (Version 2022.07.0.1; R Studio, PBC) and R (version 4.2.1; R Foundation). Training and cross-validation of the predictive algorithm were performed in Python (open-source library scikit-learn v1.1.2) [34]. Statistical significance level was defined at $p<0.05$.

Study population

Descriptive statistics for the continuous variables patient age, weight, height and BMI included mean and standard deviation (SD), while the categorical variable patient sex was reported as count and percentage. Differences in distributions between the R and MPMS group were tested by the two-sided student t-test for the continuous variables, Fischer's exact test for the variable sex and a Chi-square test-of-independence for cartilage status.

Knee morphology comparison between R and MPMS group

PC weights distributions for the first three modes of shape variation were compared between the R and MPMS group using Welch's unequal variances t-test.

Correlation analysis

Pearson correlation coefficients were calculated between the following variables: patient sex (recoded as 0 for female and 1 for male), patient age, patient length, patient weight, KOOS and the PC weights of the first three modes of shape variation for all three SSM. Statistical significant Pearson correlation coefficients were reported in a correlation matrix.

Prediction of response to treatment

The predictive value of PC weights for RTT (recoded as 1 for the R group and 0 for the MPMS group) was assessed. Prediction of RTT was performed by means of logistic regression on the PC weights in a leave-one-out cross-validation experiment. Sensitivity and specificity were calculated for detection of both response to treatment and medial post-meniscectomy syndrome [35]. In addition, the area under the curve for the receiver operating characteristic curve [35] (AUC-ROC) was calculated. A bootstrapping experiment with 1000 iterations was performed to calculate 95% confidence intervals (CIs).

14.3 Results

14.3.1 Study population

There were no significant differences between the R and MPMS group in patient sex, patient age, patient weight, patient height and body mass index (BMI) distributions (Table 12.1). Cartilage status according to the modified Outerbridge scale (intraoperative assessment) was not significantly different between the two groups for all regions. The KOOS and all of its subscales were significantly different ($p<0.001$ for all KOOS subscales, Fig. 12.4)

Table 14.1: Count and percentage for patient sex and cartilage lesion classification (modified Outerbridge), mean ± SD for variables patient age, patient weight, patient height and BMI in both R and MPMS groups.

	R group	**MPMS group**	***p* Value**
Patient sex [number (%)][a]	29 (24.2) females, 91 (75.8) males	36 (30.0) females, 84 (70.0) males	n.s.
Patient age (years)[b]	50.7 ± 12.1	52.7 ± 10.5	n.s.
Patient weight (kg)[b]	82.0 ± 15.1	82.2 ± 14.4	n.s.
Patient height (cm)[b]	177.2 ± 8.6	175.2 ± 9.6	n.s.
BMI (kg/m2)[b]	26.0 ± 3.8	26.7 ± 3.4	n.s.
Cartilage lesion classification [number (%)][c]			
Medial femoral condyle			
Grade 0–I	47 (39.2)	37 (31.4)	n.s.
Grade II	50 (41.7)	43 (36.4)	
Grade III	17 (14.1)	32 (27.1)	
Grade IV focal	6 (5.0)	6 (5.1)	
Medial tibial plateau			
Grade 0–I	56 (46.7)	57 (48.3)	n.s.
Grade II	50 (41.6)	46 (39.0)	
Grade III	11 (9.2)	13 (11.0)	
Grade IV focal	3 (2.5)	2 (1.7)	
Lateral femoral condyle			
Grade 0–I	71 (59.2)	85 (72.0)	n.s.
Grade II	47 (39.2)	29 (24.6)	
Grade III	1 (0.8)	1 (0.9)	
Grade IV focal	1 (0.8)	3 (2.5)	
Lateral tibial plateau			
Grade 0–I	70 (58.3)	85 (72.0)	n.s.
Grade II	47 (39.2)	29 (24.6)	
Grade III	1 (0.8)	2 (1.7)	
Grade IV focal	2 (1.7)	2 (1.7)	

[a] Fischer exact test.
[b] Two-sided student t test.
[c] χ^2test of independence.

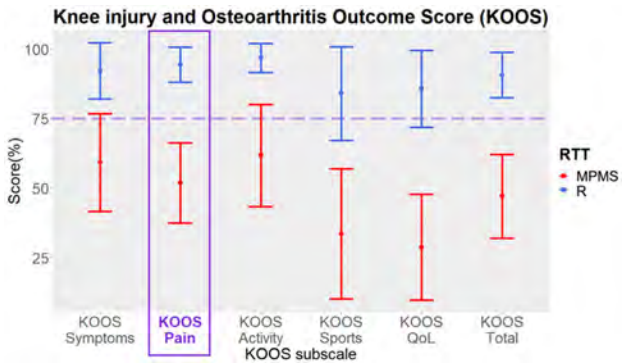


Figure 14.4: Mean ± standard deviation of all Knee injury and Osteoarthritis Outcome Score (KOOS) subscales for both medial postmeniscectomy syndrome group (MPMS) (red) and response to treatment group (R) (blue) groups. RTT, response to treatment.

14.3.2 Knee morphology comparison between R and MPMS group

Distal femur

Regarding compactness, the SSM of the distal femur captured 98.0% of the shape variance in the first 36 modes of shape variation (model compactness). Generalization error to unseen distal femur shapes in leave-one-out experiments was on average 0.30 mm (RMSE). More detailed performance metrics of the statistical shape models can be found in the Supplementary Information. The first mode of shape variation is the three-dimensional size of the distal femur (Fig. 12.5, left column). Distal femora of the MPMS group were significantly smaller than those in the R group ($p<0.001$). The second mode of shape variation (Fig. 12.5, middle column) captured the mediolateral intercondylar notch width. The MPMS group had a significantly wider intercondylar notch ($p<0.001$). The third mode of shape (Fig. 12.5, right column) variation encompassed the mediolateral (ML) width of the medial femoral condyle and anteroposterior (AP) length of both femoral condyles. MPMS knees had a significantly smaller ML medial femoral condyle and larger AP femoral condyles ($p<0.001$). Together, these three modes of shape variation accounted for 86.7% of the total femoral shape variance.

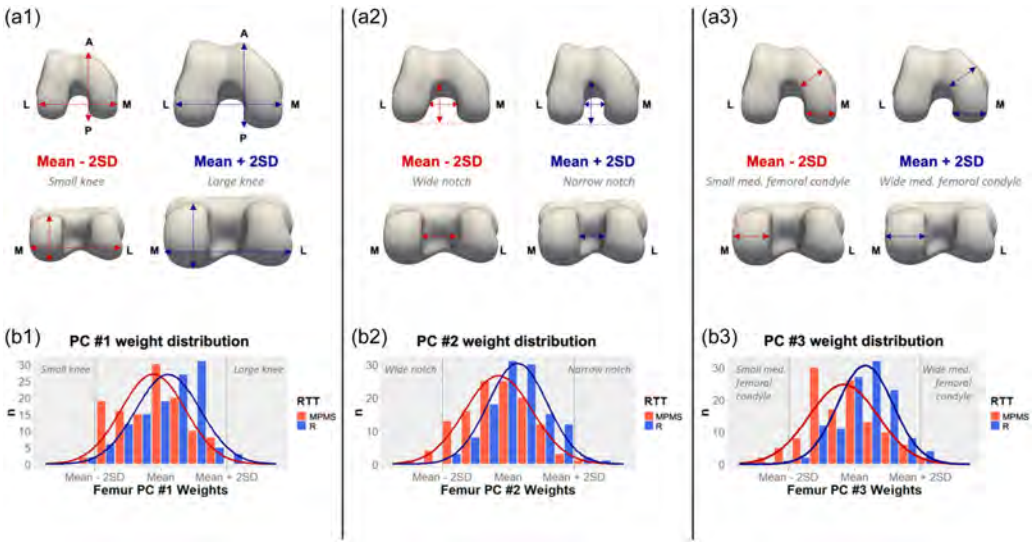


Figure 14.5: (a1–a3) First three modes of shape variation in distal femur: (top) inferior view, (bottom) posterior view. **(b1–b3)** Histogram of principal component (PC) weights distribution for the first three modes of shape variation in medial postmeniscectomy syndrome group (red) and response to treatment group (R) (blue) group. A, anterior; P, posterior; L, lateral; M, medial; MPMS, medial postmeniscectomy syndrome group; n, number of cases; RTT, response to treatment; SD, standard deviations.

Proximal tibia

The SSM of the proximal tibia described 98.1% of the shape variance in the first 37 modes of shape variation that. Leave-one-out experiments resulted in an average generalizability error of 0.25 mm RMSE. The first mode of shape variation (Fig. 12.6, left column) in the proximal tibia was three-dimensional size. The MPMS group consisted of significantly smaller tibial plateaus (p=0.005). The second mode of shape variation (Fig. 12.6, middle column) included mediolateral width of the tibial plateau and interspine distance, as well as tibial spine height variation. The MPMS group had a significantly wider interspine distance and lower tibial spines (p<0.001). The third mode of shape variation (Fig. 12.6, right column) included the AP length of the tibial plateau and the sagittal concavity of the medial tibial plateau. A small difference was observed between the R and MPMS groups for this mode of shape variation (p=0.01), where tibial plateaus from the MPMS group had a more pronounced mediolateral width relative to their anteroposterior depth and a less concave (sagittal plane) medial tibial plateau. Together, these three modes of shape variation explained 83.7% of the total tibial shape variance.

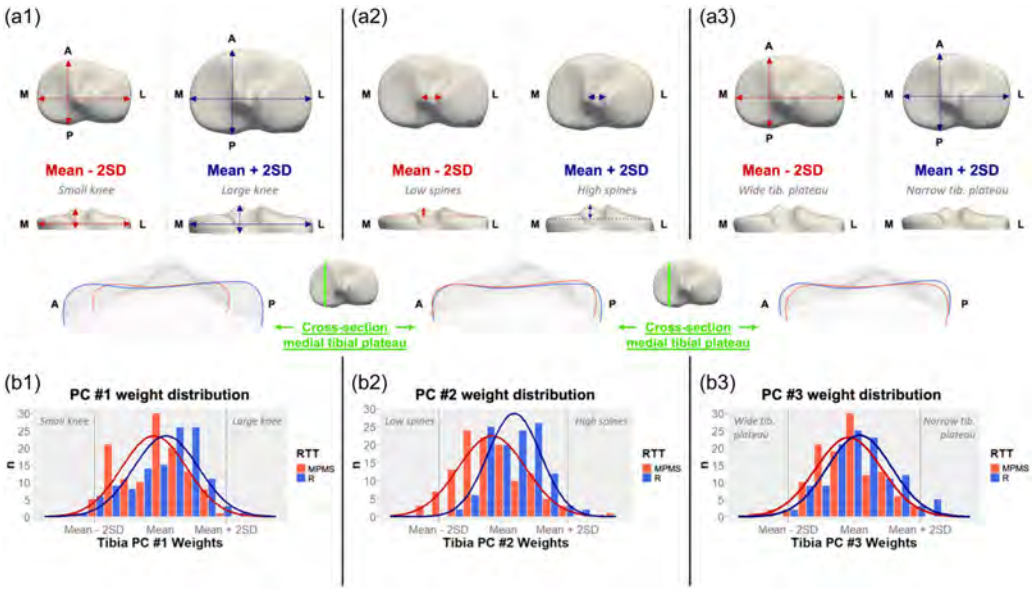


Figure 14.6: (a1–a3) First three modes of shape variation in proximal tibia: (top) superior view, (middle) posterior view, (bottom) crosssectional medial view of the medial tibial plateau. **(b1–b3)** Histogram of principal component (PC) weights distribution for the first three modes of shape variation in medial postmeniscectomy syndrome group (MPMS) (red) and response to treatment group (R) (blue) group. A, anterior; L, lateral, M, medial; n, number of cases, P, posterior; RTT, response to treatment; SD, standard deviations.

Tibiofemoral joint (combined shape model in neutral position)

The combined SSM of femur and tibia together captured 98.0% of the variance in 46 modes of shape variation as a measure of the model compactness. This corresponded to a generalization error of 0.34 mm in the leave-one-out experiments. The main mode of shape variation (Fig. 12.7, left column) in the tibiofemoral joint was size and accounted for 83.4% of the total shape variance. The distribution of the PC weights was significantly shifted towards smaller knees in the MPMS group ($p<0.001$). The second mode of shape variation (Fig. 12.7, middle column) described the intercondylar mediolateral notch width and tibial interspine distance. A wider femoral intercondylar notch and larger tibial interspine distance was observed in the MPMS group ($p<0.001$). The third mode of shape variation (Fig. 12.7, right column) included the mediolateral width of the medial femoral condyle and the height of the tibial spines. A significantly smaller medial femoral condyle and less pronounced tibial spines were observed in the MPMS knees ($p<0.001$). Together, these three modes of shape variation described 88.2% of the total tibiofemoral shape variance.

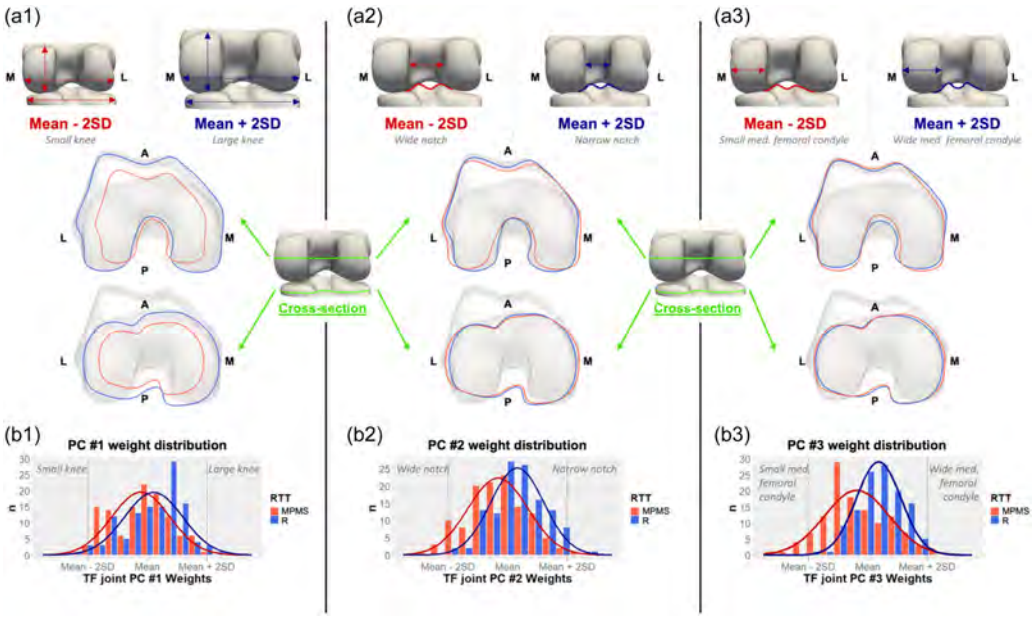
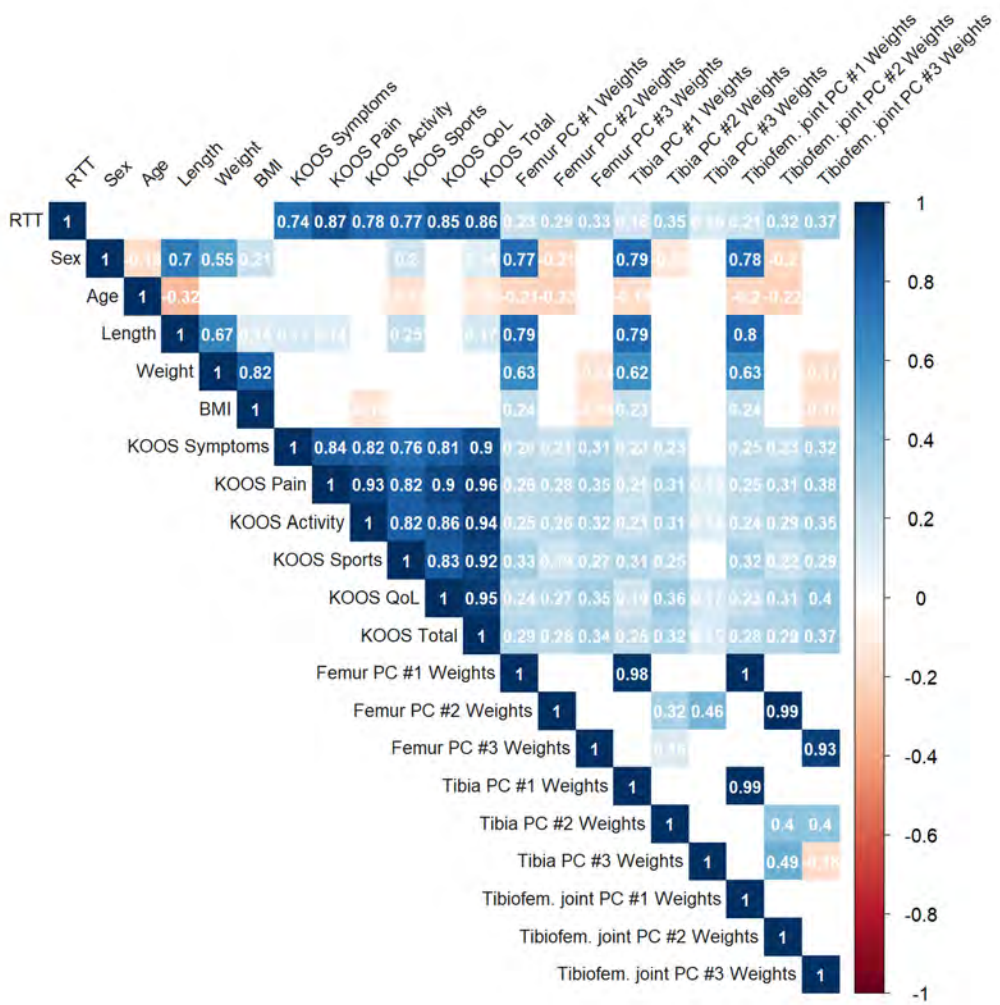


Figure 14.7: (a1–a3) First three modes of shape variation in tibiofemoral joint: (top) posterior view, (middle) inferior view cross-sections of distal femur, (bottom) inferior view cross-sections of proximal tibia. **(b1–b3)** Histogram of principal component (PC) weights distribution for the first three modes of shape variation in medial postmeniscectomy syndrome group (MPMS) (red) and response to treatment group (R) (blue) group. A, anterior; L, lateral, M, medial; n, number of cases; P, posterior; RTT, response to treatment; SD, standard deviations.

14.3.3 Correlation analysis between demographic, clinical and knee morphology variables

Statistical significant Pearson correlation coefficients (R^2) were summarized in a correlation matrix (Table 12.2) for the demographic, clinical and morphological variables of the distal femur, proximal tibia and tibiofemoral joint. RTT was encoded as 0 for the MPMS group and 1 for the R group. RTT was strongly positively correlated with all KOOS subscores ($p<0.001$). For all three SSM's (femur, tibia and tibiofemoral joint), the PC weights of the first three modes of shape variation were all very weakly to weakly (R^2 between 0.16 and 0.37) correlated with RTT ($p\leq0.01$). None of the demographic or clinical variables were correlated with RTT. Patient sex, length and weight were strongly positively correlated (R^2 between 0.62 and 0.80) with the first mode of shape variation (three-dimensional size) for all three SSM's ($p<0.001$). Patient sex showed a very weak and weak negative correlation with the second mode of shape variation (femoral notch width and tibial interspine distance) from respectively the SSM of the proximal tibia (R2=-0.13, p=0.047) and the SSM of distal femur and tibiofemoral joint (R2 between -0.21 and 0.20, $p < 0.002$).

Table 14.2: Correlation matrix of demographic, clinical, PROM- and SSM-derived variables.



14.3.4 Prediction of response to treatment based on knee morphology

Distal femur morphology is an independent predictor for RTT. Solely based on the distal femur PC weights, MPMS was predicted with a sensitivity of 74.8% (95% CI 74.3% to 75.3%) and a specificity of 80.3% (95% CI 79.8% to 80.8%) Area under the receiver operating characteristic curve (AUC-ROC, Figure 12.8, yellow curve) for this classifier was 0.827 (95% CI 0.824 to 0.830). The predictive logistic regression algorithm with proximal tibia PC weights as input, identified MPMS in a leave-one-out cross-validation experiment with a sensitivity of 74.7% (95% CI 74.2% to 75.2%) and a specificity of 78.5% (95% CI 78.0% to 79.0%). AUC for the ROC curve (Figure 12.8, blue curve) was 0.836 (95% CI 0.833 to 0.839) for this classifier based on proximal tibia

morphology. Tibiofemoral joint morphology predicted MPMS with a sensitivity of 74.9% (95% CI 74.4% to 75.4%) and a specificity of 81.0% (95% CI 80.6% to 81.5%) in a leave-one-out cross-validation experiment. AUC-ROC was 0.840 (95% CI 0.837 to 0.843) for this tibiofemoral joint morphology based classifier. The receiver operating characteristic curve for the classifier using a set of tibiofemoral joint shape features is shown in Figure 12.8 (green curve).

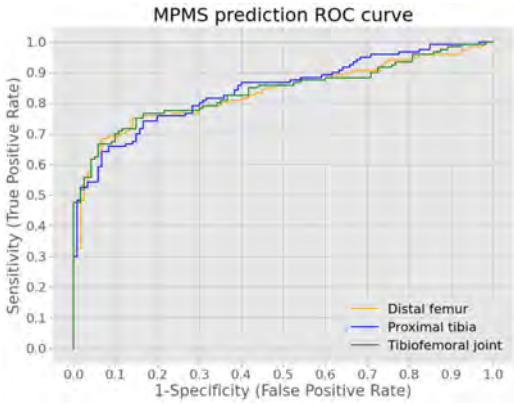


Figure 14.8: Receiver operating characteristic curves for three distinct classifiers, using a set of distal femur, proximal tibia or tibiofemoral joint shape features as predictor for medial postmeniscectomy syndrome (MPMS). ROC, receiver operating characteristic.

14.4 Discussion

This study unveils for the first time bony knee morphological differences between responders to APMM and medial post-meniscectomy syndrome patients: a smaller overall size of the knee, a wider intercondylar notch and a smaller medial femoral condyle were the main morphological variations identified in medial post-meniscectomy syndrome knees. Moreover, as a second key finding in this study, morphology-based predictive models demonstrated a sensitivity of more than 75% and a specificity surpassing 80% in anticipating the outcome of APMM. Based on these findings, a biomechanical mechanism is hypothesized to explain the main differentiating mode of shape variation between responders to APMM and medial post-meniscectomy syndrome. While often excluded or filtered out in statistical shape analysis, a significant difference in knee size was observed between the R and MPMS group knees for the three SSMs: isolated femur, isolated tibia and combined in the tibiofemoral joint. The knees from the MPMS group patients were significantly smaller than their counterparts in the R group, in contrast with no differences in patient length, weight and BMI between both groups. A smaller knee implicates a smaller articular contact surface area. In biomechanics, pressure is defined as force per unit area of surface. Therefore, a smaller knee loaded with a similar body weight is subject to higher pressure. Different studies already established increased contact pressure as a central driver in progressive knee degeneration [36, 37].

To the authors' knowledge, this is the largest 3D knee morphology database of

post-meniscectomized patients, consisting of imaging, demographic and clinical (incl. PROM) data at 2 years post-operatively. Similar techniques based on statistical shape modeling have already been applied on other pathologies, e.g. for automatic staging of trochlear dysplasia [22] or identifying morphological bone variations linked to scaphoid fractures [38]. Previously, our pilot study on medial compartment degeneration already identified the small medial femoral condyle morphotype [16] and its potential association with early medial compartment degeneration. Several in-silico and in-vitro studies have confirmed the load- distributing function of the meniscus, as well as the increase in cartilage peak stress after partial meniscectomy [39]. The smaller contact surface area in the small medial femoral condyle morphotype could already imply a higher peak stress compared to a wide medial femoral condyle knee. By further reducing the contact surface area, meniscus functional loss by degeneration or APMM might additionally increase that peak stress, leading to insufficiency of the cartilage and subchondral bone to bear the load [40]. This is clinically reflected in the post-meniscectomy syndrome. Several studies established the association between coronal alignment of the knee and pathology [41, 42]. In our study, the influence of alignment as a latent prognostic factor was minimized since all knees had a coronal alignment within physiological range per clinical judgement. The combination of a full lower-limb phenotype analysis with this confined knee morphology analysis might even result in more robust algorithms to predict APMM outcome, irrespective of knee malalignment. A clustering analysis by Hohlmann et al. [43] revealed no distinct morphotypes in end-stage knee OA knees. In contrast to their conclusion, our current study found that knee shape (including size and aspect ratio) is highly correlated with patient demographics (sex, length, weight) and even APMM outcome. These contradictory findings can be explained by the fact that unsupervised clustering algorithms are typically less powerful for classification purposes. Importantly, the previous authors corrected for both size and aspect ratio, while these parameters were found to be significantly correlated to several clinical variables in the current study. Based on statistical shape models, Bowes et al. created a score to quantify radiological disease progression in cases from the Osteoarthritis Initiative (OAI) database [44, 45]. Tack et al. [46] extended this concept further to a set of osteoarthritis biomarkers by adding traditional measurements (e.g. volume and surface area) and SSM-derived features for femoral bone, tibial bone, medial and lateral meniscus. This set of biomarkers proved superior performance for prediction of total knee replacement within one year in comparison with the bony SSM-derived featur es alone. The aforementioned studies support the hypothesis of bone shape as a biomarker and predictor for wear-related pathology. Indeed, the post-meniscectomy syndrome also lies in the spectrum of degenerative knee pathologies and this current study was able to predict its onset solely based on the knee morphology.

The relevance of this work lies potentially in the health-economical aspect of medial meniscus lesion treatment [47, 11]. This is a promising first step towards a better identification of candidate patients for APMM [48]. Smarter risk definitions and patient selection in the future will result in a more personalized health care [49]. Eventually, forthcoming clinical decision support systems might serve as a tool for clinicians to help achieving this aim.

The strengths of this study include the multi-center study design, the high volume of manually segmented MRI scans, as well as the unbiased methodology to describe knee morphology quantitatively. Instead of using predefined measurements, patterns of

morphological variation were extracted by a robust mathematical algorithm. Furthermore, this high level of automated analysis starting from the 3D models of femur and tibia allows the analysis of even larger databases at a low marginal time cost (even zero human marginal time cost). Another strength of this unique dataset is the highly homogenous group of patients, and the equal distribution of collected potential latent demographic variables (patient sex, height and weight) influencing knee morphology across both R and MPMS groups. Finally, the pre-operative imaging used to create the 3D models is already available for this pathology, even in a standard clinical setting.

This study shows promising results, despite some limitations. Firstly, because of the multi-center retrospective study design, no precise data was available on possibly confounding variables such as personalized medication and rehabilitation protocol. However, the clinical centers (in Antwerp, Milan and Regensburg) have closely matched patient demographics and healthcare systems. All patients received nearly identical post-operative care, adhering to the latest clinical guidelines. Secondly, during the follow-up period after APMM, no new MRI scans were performed unless clinically necessary. As a result, quantitative standardized measurements of the resected meniscal tissue could not be calculated, nor retrieved from the surgery report and no subgroup analysis based on the presence of radiologic signs such as bone marrow edema, osteonecrosis or meniscal retear could be performed. Finally, it should be noted that only patients with a clinically normal coronal malalignment were included in this study, per surgeon clinical assessment. In a real world clinical protocol for meniscus lesion management, no full-leg standing radiographs, nor full-leg CT scans, are indicated for these patients. Given the retrospective nature of this study, exact coronal alignment measurements were therefore unavailable. Future research includes large prospective validation studies, with additional follow-up imaging, to warrant the generalizability of the findings in this study.

14.5 Conclusion

In conclusion, morphological variations determine the clinical outcome of APMM in a patient population with comparable weight, height, alignment and cartilage status. More specifically, a smaller total knee size, a wider intercondylar notch and a smaller medial femoral condyle are strongly linked with post-meniscectomy syndrome. In addition, a predictive model was able to anticipate the clinical outcome of APMM at two years of follow-up with a sensitivity above 75% and a specificity exceeding 80%.

Bibliography

[1] Beaufils P, Becker R, Kopf S, Matthieu O, Pujol N. The knee meniscus: management of traumatic tears and degenerative lesions. EFORT open reviews. 2017 5;2:195-203.

[2] Hall MJ, Schwartzman A, Zhang J, Liu X. Ambulatory Surgery Data From Hospitals and Ambulatory Surgery Centers: United States, 2010. National health statistics reports. 2017 2:1-15.

[3] Jacquet C, Pujol N, Pauly V, Beaufils P, Ollivier M. Analysis of the trends in arthroscopic meniscectomy and meniscus repair procedures in France from 2005 to 2017. Orthopaedics & Traumatology: Surgery & Research. 2019 6;105:677-82.

[4] Katano H, Koga H, Ozeki N, Otabe K, Mizuno M, Tomita M, et al. Trends in isolated meniscus repair and meniscectomy in Japan, 2011–2016. Journal of Orthopaedic Science. 2018 7;23:676-81.

[5] Thorlund JB, Hare KB, Lohmander LS. Large increase in arthroscopic meniscus surgery in the middle-aged and older population in Denmark from 2000 to 2011. Acta orthopaedica. 2014 6;85:287-92.

[6] Reito A, Harris IA, Karjalainen T. Arthroscopic partial meniscectomy: did it ever work? Acta orthopaedica. 2021 10;93:1-10.

[7] Abrams GD, Frank RM, Gupta AK, Harris JD, McCormick FM, Cole BJ. Trends in Meniscus Repair and Meniscectomy in the United States, 2005-2011. The American Journal of Sports Medicine. 2013 10;41:2333-9.

[8] Beaufils P, Becker R, Kopf S, Englund M, Verdonk R, Ollivier M, et al. Surgical management of degenerative meniscus lesions: the 2016 ESSKA meniscus consensus. Knee Surgery, Sports Traumatology, Arthroscopy. 2017 2;25:335-46.

[9] Kopf S, Beaufils P, Hirschmann MT, Rotigliano N, Ollivier M, Pereira H, et al. Management of traumatic meniscus tears: the 2019 ESSKA meniscus consensus. Knee Surgery, Sports Traumatology, Arthroscopy. 2020 4;28:1177-94.

[10] Drobnič M, Ercin E, Gamelas J, Papacostas ET, Slynarski K, Zdanowicz U, et al. Treatment options for the symptomatic post-meniscectomy knee. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 6;27:1817-24.

[11] Zaslav KR, Farr J, Alfred R, Alley RM, Dyle M, Gomoll AH, et al. Treatment of post-meniscectomy knee symptoms with medial meniscus replacement results in greater pain reduction and functional improvement than non-surgical care. Knee Surgery, Sports Traumatology, Arthroscopy. 2022 4;30:1325-35.

[12] Rao AJ, Erickson BJ, Cvetanovich GL, Yanke AB, Bach BR, Cole BJ. The Meniscus-Deficient Knee. Orthopaedic Journal of Sports Medicine. 2015 10;3:232596711561138.

[13] Salata MJ, Gibbs AE, Sekiya JK. A Systematic Review of Clinical Outcomes in Patients Undergoing Meniscectomy. The American Journal of Sports Medicine. 2010 9;38:1907-16.

[14] Willinger L, Lang JJ, Berthold D, Muench LN, Achtnich A, Forkel P, et al. Varus alignment aggravates tibiofemoral contact pressure rise after sequential medial meniscus resection. Knee Surgery, Sports Traumatology, Arthroscopy. 2020 4;28:1055-63.

[15] Eijgenraam SM, Reijman M, Bierma-Zeinstra SMA, Yperen DTV, Meuffels DE. Can we predict the clinical outcome of arthroscopic partial meniscectomy? A systematic review. British Journal of Sports Medicine. 2018 4;52:514-21.

[16] Grammens J, Haver AV, Danckaers F, Booth B, Sijbers J, Verdonk P. Small medial femoral condyle morphotype is associated with medial compartment degeneration and distinct morphological characteristics: a comparative pilot study. Knee Surgery, Sports Traumatology, Arthroscopy. 2021 6;29:1777-89.

[17] Tabbaa SM, Pace JL, Frank RM, Grammens J, Verdonk P. Meniscus Size Differs Between Patient and Donor Populations for Meniscus Allograft Transplantation. Arthroscopy, sports medicine, and rehabilitation. 2023 6;5:e569-76.

[18] Fernández-Jaén T, López-Alcorocho JM, Rodriguez-Iñigo E, Castellán F, Hernández JC, Guillén-García P. The Importance of the Intercondylar Notch in Anterior Cruciate Ligament Tears. Orthopaedic Journal of Sports Medicine. 2015 8;3:2325967115597882.

[19] Haver AV, Roo KD, Beule MD, Labey L, Baets PD, Dejour D, et al. The Effect of Trochlear Dysplasia on Patellofemoral Biomechanics. The American Journal of Sports Medicine. 2015 6;43:1354-61.

[20] Polamalu SK, Novaretti J, Musahl V, Debski RE. Tibiofemoral bony morphology impacts the knee kinematics after anterolateral capsule injury and lateral extraarticular tenodesis differently than intact state. Journal of Biomechanics. 2022 6;139:110857.

[21] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models-Their Training and Application. Comput Vis Image Underst. 1995 1;61:38-59.

[22] Cerveri P, Belfatto A, Baroni G, Manzotti A. Stacked sparse autoencoder networks and statistical shape models for automatic staging of distal femur trochlear dysplasia. Int J Med Robot. 2018 12;14:e1947.

[23] Haver AV, Mahieu P, Claessens T, Li H, Pattyn C, Verdonk P, et al. A statistical shape model of trochlear dysplasia of the knee. The Knee. 2014 3;21:518-23.

[24] Meynen A, Matthews H, Nauwelaers N, Claes P, Mulier M, Scheys L. Accurate reconstructions of pelvic defects and discontinuities using statistical shape models. Computer methods in biomechanics and biomedical engineering. 2020 10;23:1026-33.

[25] Osstyn J, Danckaers F, Haver AV, Oramas J, Vanhees M, Sijbers J. Automated Virtual Reduction of Displaced Distal Radius Fractures. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). IEEE; 2023. p. 1-4.

[26] Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health and Quality of Life Outcomes. 2003;1:64.

[27] Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS) - Development of a self-administered outcome measure. J Orthop Sports Phys Ther. 1998 8;28:88-96.

[28] Agarwalla A, Gowd AK, Liu JN, Lalehzarian SP, Christian DR, Cole BJ, et al. Predictive Factors and Duration to Return to Sport After Isolated Meniscectomy. Orthopaedic Journal of Sports Medicine. 2019 4;7:232596711983794.

[29] Audenaert EA, Pattyn C, Steenackers G, Roeck JD, Vandermeulen D, Claes P. Statistical Shape Modeling of Skeletal Anatomy for Sex Discrimination: Their Training Size, Sexual Dimorphism, and Asymmetry. Frontiers in Bioengineering and Biotechnology. 2019 11;7:302.

[30] Asano H, Muneta T, Ikeda H, Yagishita K, Kurihara Y, Sekiya I. Arthroscopic evaluation of the articular cartilage after anterior cruciate ligament reconstruction: a short-term prospective study of 105 patients. Arthroscopy. 2004 5;20:474-81.

[31] Danckaers F, Huysmans T, Lacko D, Ledda A, Verwulgent S, Dongen SV, et al. Correspondence Preserving Elastic Surface Registration with Shape Model Prior. In: 2014 22nd International Conference on Pattern Recognition. IEEE; 2014. p. 2143-8.

[32] Audenaert EA, den Eynde JV, de Almeida DF, Steenackers G, Vandermeulen D, Claes P. Separating positional noise from neutral alignment in multicomponent statistical shape models. Bone Reports. 2020 6;12:100243.

[33] Goparaju A, Csecs I, Morris A, Kholmovski E, Marrouche N, Whitaker R, et al. On the Evaluation and Validation of Off-the-Shelf Statistical Shape Modeling Tools: A Clinical Application. In: Reuter M, Wachinger C, Lombaert H, Paniagua B, Lüthi M, Egger B, editors. Shape in Medical Imaging. Springer International Publishing; 2018. p. 14-27.

[34] Pedregosa F, Varoquaux G, Gramfort A, V BM, Thirion, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-30.

[35] Riffenburgh RH, Gillen DL. In: Risks, odds, and receiver operating characteristic curves. Elsevier; 2020. p. 215-38.

[36] Mononen ME, Tanska P, Isaksson H, Korhonen RK. New algorithm for simulation of proteoglycan loss and collagen degeneration in the knee joint: Data from the osteoarthritis initiative. J Orthop Res. 2018 6;36:1673-83.

[37] Pattappa G, Zellner J, Johnstone B, Docheva D, Angele P. Cells under pressure - the relationship between hydrostatic pressure and mesenchymal stem cell chondrogenesis. European cells & materials. 2019 5;37:360-81.

[38] Bevers MSAM, Wyers CE, Daniels AM, Audenaert EA, van Kuijk SMJ, van Rietbergen B, et al. Association between bone shape and the presence of a fracture in patients with a clinically suspected scaphoid fracture. Journal of Biomechanics. 2021 11;128:110726.

[39] Sukopp M, Schall F, Hacker SP, Ignatius A, Dürselen L, Seitz AM. Influence of Menisci on Tibiofemoral Contact Mechanics in Human Knees: A Systematic Review. Frontiers in bioengineering and biotechnology. 2021;9:765596.

[40] Lee SJ, Aadalen KJ, Malaviya P, Lorenz EP, Hayden JK, Farr J, et al. Tibiofemoral Contact Mechanics after Serial Medial Meniscectomies in the Human Cadaveric Knee. The American Journal of Sports Medicine. 2006 8;34:1334-44.

[41] Hirschmann MT, Moser LB, Amsler F, Behrend H, Leclerq V, Hess S. Functional knee phenotypes: a novel classification for phenotyping the coronal lower limb alignment based on the native alignment in young non-osteoarthritic patients. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 5;27:1394-402.

[42] Oevelen AV, den Borre IV, Duquesne K, Pizurica A, Victor J, Nauwelaers N, et al. Wear patterns in knee OA correlate with native limb geometry. Frontiers in bioengineering and biotechnology. 2022;10:1042441.

[43] Hohlmann B, Asseln M, Xu J, Radermacher K. Investigation of morphotypes of the knee using cluster analysis. The Knee. 2022 3;35:157-63.

[44] Bowes MA, Kacena K, Alabas OA, Brett AD, Dube B, Bodick N, et al. Machine-learning, MRI bone shape and important clinical outcomes in osteoarthritis: data from the Osteoarthritis Initiative. Annals of the Rheumatic Diseases. 2021 4;80:502-8.

[45] Lester G. Clinical research in OA - The NIH Osteoarthritis Initiative. Journal of musculoskeletal & neuronal interactions. 2008;8:313-4.

[46] Tack A, Ambellan F, Zachow S. Towards novel osteoarthritis biomarkers: Multi-criteria evaluation of 46,996 segmented knee MRI data from the Osteoarthritis Initiative. PloS one. 2021;16:e0258855.

[47] Hershman EB, Jarvis JL, Mick T, Dushaj K, Elsner JJ. Direct treatment cost outcomes among patients with medial meniscus deficiency: results from a 24-month surveillance study. Current Medical Research and Opinion. 2020 3;36:427-37.

[48] Pihl K, Ensor J, Peat G, Englund M, Lohmander S, Jørgensen U, et al. Wild goose chase – no predictable patient subgroups benefit from meniscal surgery: patient-reported outcomes of 641 patients 1 year after surgery. British Journal of Sports Medicine. 2020 1;54:13-22.

[49] Vicente AM, Ballensiefen W, Jönsson JI. How personalized medicine will transform healthcare by 2030: the ICPerMed vision. Journal of Translational Medicine. 2020 12;18:180.

Chapter 15

Machine learning for APMM outcome prediction: clinical and morphological data from the MEFISTO project

This chapter describes the final aim of this research project: a predictive model to stratify between responders to arthroscopic partial medial meniscectomy (APMM) and medial post-meniscectomy syndrome (MPMS) patients, based on a combination of bony morphological features (derived from pre-operative imaging) and a confined set of clinical features.

> This chapter is adopted from a manuscript submitted at Computer Methods and Programs in Biomedicine:
> *Grammens, J., Danckaers, F., Van Haver, A., Verdonk, P., Sijbers, J. Shaping predictive models for arthroscopic partial medial meniscectomy outcome: a machine learning approach using clinical and bony morphological data from the MEFISTO Project.*

15.1 Introduction

Arthroscopic partial medial meniscectomy (APMM) is one of the most frequently performed orthopedic procedures worldwide [1, 2, 3, 4, 5], particularly for treating meniscal tears that lead to pain, mechanical symptoms, and functional limitations in the knee joint [6]. Meniscal injuries can significantly disrupt the biomechanics of the knee joint by altering load distribution, impacting stability, and contributing to degenerative changes over time [7, 8]. APMM, which involves the partial resection of the damaged or loose meniscal tissue, can alleviate symptoms and improve function in the short term [6]. However, long-term outcomes after APMM vary widely, with a subgroup of patients experiencing progressive osteoarthritic changes, pain recurrence, and functional limitations [9]. As such, the prediction of post-operative outcomes has gained interest to perform a better, data-driven, patient selection.

The variability in APMM outcomes is influenced by a range of patient-specific factors, including age, body mass index (BMI), activity level, and the characteristics of the

meniscal tear itself (e.g., tear location, pattern, and chronicity). Furthermore, the health of the knee joint at the time of surgery, particularly the integrity of the cartilage, plays a critical role in determining long-term recovery and function. Studies have shown that patients with pre-existing cartilage damage or early signs of osteoarthritis tend to experience less favorable outcomes after APMM, with an increased risk of subsequent joint degeneration [10, 11]. MRI imaging, which provides detailed visualizations of both soft tissue and cartilage, has become a valuable tool for assessing these risk factors pre-operatively, as it allows for precise evaluation of the meniscus, cartilage, and surrounding joint structures. However, while individual clinical and imaging variables have been associated with APMM outcomes, integrating these variables into a cohesive, accurate prognostic model remains a challenge [12].

At the severe end of the post-surgical degenerative spectrum is the medial post-meniscectomy syndrome (MPMS), a debilitating condition characterized by a dull and nagging knee pain after a short pain-free interval, affecting between 6 and 25% of patients who undergo APMM [13]. The existing literature on risk factors for this complication is largely based on generalized factors associated with knee degeneration and osteoarthritis, which complicates the ability to predict whether a patient will experience a rather slow to moderate progression of knee osteoarthritis or a really rapid degeneration combined with severe inflammatory symptoms, such as seen in the medial post-meniscectomy syndrome. Unfortunately, alternative treatments such as meniscus preservation or substitution are associated with scarce supply or higher failure rates in the short- to intermediate-term compared to partial medial meniscectomy, thereby limiting their viability as first-line treatments [14]. Consequently, APMM remains the standard treatment for most patients.

Despite the currently high failure rates of meniscus substitution options [15, 16], further and more extensive development of these alternative treatments to partial meniscectomy may not be immediately perceived as a priority, given the already high (but short-term) success rates of the meniscectomy as first-line treatment. Although post-meniscectomy syndrome affects only approximately 10% of patients [13], the total number of individuals impacted is significant, given that meniscectomy is one of the most frequently performed orthopedic surgeries globally. The potential for long-term disability and the considerable healthcare burden associated with a meniscus-deficient knee underscores the need for more effective treatment strategies. The ability to predict the likelihood of developing the medial post-meniscectomy syndrome preoperatively could suggest a more favorable outcome for alternative treatments in a significant number of patients.

In recent years, machine learning (ML) has offered a promising research topic for developing prognostic models that can incorporate and analyze complex, multidimensional datasets. ML techniques have been successfully used to predict complications in other orthopedic surgeries, such as total knee arthroplasty (TKA), where models that account for patient demographics and clinical parameters have been shown to achieve a high prognostic accuracy and support personalized treatment decisions [17] However, similar prognostic tools specific to APMM are still lacking, despite the variation in APMM outcomes and the potential value of a predictive tool to guide surgical decisions. The purpose of this study is to develop a data-driven prognostic model for APMM outcomes, using a combination of clinical, demographic, and imaging data to estimate the risk of developing medial post-meniscectomy syndrome.

An additional factor that may influence APMM outcomes is the morphology of the knee's bony structures, which can impact knee kinematics and biomechanics and potentially exacerbate the effects of meniscal injury [18, 19]. Bony morphology characteristics, such as the tibial plateau slope, the shape of the femoral condyles, and the alignment of the lower limb, have been linked to joint mechanics and load distribution patterns that affect the meniscus and cartilage [20]. For instance, an increased tibial slope has been associated with higher shear forces on the meniscus, which may predispose individuals to meniscal tears and influence recovery post-meniscectomy [21]. Similarly, varus or valgus knee alignment has been implicated in altering the medial or lateral compartment loading, potentially affecting healing and progression of joint degeneration [22]. By including bony morphology characteristics in a prognostic model, it may be possible to gain a more comprehensive understanding of the biomechanical environment of the knee [23], thereby improving predictive accuracy for APMM outcomes and allowing for more personalized surgical recommendations.

This research is part of the MEFISTO project, where two alternative meniscus substitution devices were developed: a bioactive resorbable meniscal scaffold [24, 25] and a bioactive unloading prosthesis (artificial meniscus implant)[26] for meniscus substitution. In the context of these innovations, an accurate predictive algorithm for APMM was required, more specific to predict the onset of the MPMS. Thus, the primary objective of this study is to estimate the risk of developing medial post-meniscectomy syndrome pre-operatively. This will enable the identification of optimal candidates for these alternative meniscus substitution therapies as opposed to those who may already benefit from the standard APMM. In this paper, we describe the development and validation of a prognostic model for APMM that integrates clinical, demographic, and bony morphological data. It represents a first step toward a more personalized approach to the treatment of meniscus injuries. By reducing the incidence of MPMS through avoiding ineffective meniscectomy procedures in a select number of patients, patient outcomes could be significantly improved, while also reducing the long-term socioeconomic impact associated with the meniscus-deficient knee.

15.2 Materials and methods

15.2.1 Study design and data collection

A retrospective study was designed, using post-meniscectomy patient data from three high-volume orthopaedic centers (Antwerp, Milan and Regensburg) [27]. It was approved by each site's Ethics Committee and was performed according to the principles of the declaration of Helsinki. All patients provided informed consent prior to their inclusion in the study. Outcome stratification for response to treatment (R group) versus post-meniscectomy syndrome (MPMS group; treatment failure) was done by means of the KOOS pain score (> 75: R group, < 75: MPMS group) or revision surgery to the index knee within 6 months (MPMS group) after the primary APMM. Following a priori power analysis, predefined sample sizes were set at 120 patients per group. Patient inclusion to each group was ceased upon reaching the predefined sample size for that group and attempts to contact patients with incomplete KOOS questionnaire or informed consent were stopped when both groups were complete

(n=240).

Inclusion/exclusion criteria

Patients were eligible for the study if they were between 18 and 70 years old, had a primary medial meniscus lesion, for which meniscal repair or conservative treatment was not applicable and hence an APMM was indicated and performed by an expert surgeon. Predefined exclusion criteria were unavailable pre-operative MRI, inability to communicate, an unstable knee (IKDC grade C or D), patellar instability or trochlear dysplasia, limited knee range of motion (IKDC grade C or D), cartilage lesions (grade IV and larger than 2cm, non-focal), coronal malalignment (as judged clinically), concomitant discoid meniscus, morbid obesity (BMI > 35), a history of meniscus repair or major lower limb surgery prior to the meniscectomy, septic or rheumatoid arthritis, neurological disorders, posterior cruciate ligament repair or reconstruction, insufficiency fractures or avascular necrosis, plica syndrome or less than 2mm intact medial meniscal rim left intra-operatively.

Collected data

For all patients, the following data was collected: demographic and clinical data from the patient file and the pre-operative MRI scan. The collected demographic data included patient age and biological sex. In addition, the clinical data was comprised of patient weight, length, and concomittant involvement of a lateral meniscus lesion. BMI was calculated as measure of overweight from patient weight and length, using the following formula:

$$\text{BMI} = \frac{\text{weight[kg]}}{(\text{length[m]})^2} \tag{15.1}$$

Consequently BMI categories were defined based on the following thresholds: lower than 25, between 25 and 27.5, between 27.5 and 30 and more than 30 kg/m^2. Patient reported outcome measures (PROM's) were collected from the KOOS questionnaire at 2 years of follow-up and served as measure to assess the response to treatment.

Pre-processing of imaging data

Pre-operative MRI scans were loaded as DICOM files into the segmentation software (Mimics 23.0, Materialise NV, Leuven, Belgium) to create patient-specific 3D bone models. Sequences in the three perpendicular anatomical planes were available, following the standard knee scan protocols. Distal femur and proximal tibia were segmented into two distinct 3D models, consisting of the bone and cartilage layer. The resulting 3D models were projected on all available sequences and adjusted manually using the "Contour edit" tool. Finally, the 3D models of distal femur and proximal tibia were saved as triangular meshes and further used as input for the morphometric characterization.

15.2.2 Knee shape characterization: surface registration for statistical shape analysis

Registration of the bone meshes was performed following the algorithm described by Danckaers et al [28]. As a result, a set of dense anatomically corresponding bone meshes is obtained. That is, all mesh surface points are ordered in a homologous way to ensure anatomical correspondence. Next, following subtraction of the mean bone shape, principal component analysis (PCA) extracted the principal components or main modes of shape variation per bone shape. Each bone mesh can now be reconstructed by the mean bone mesh plus a linearly weighted (by the principal component scores) combination of these principal components. As these modes of shape variation were highly correlated between femur and tibia (e.g. large femur on large tibia), an additional PCA step was performed on the united shape modes of femur and tibia (concatenated). That way, the modes of bony shape variation over the tibiofemoral joint are calculated, not taking into account the relative position of femur with respect to tibia. As PCA is a purely linear operation, all steps involving PCA can be inverted for matters of explainability and visualizations. Of note, the construction of the statistical shape model was embedded into the scikit-learn machine learning pipeline (as described in the next section), ensuring consistent data transformation while safeguarding the strict separation of training, validation and test data of the chosen nested cross-validation approach. This way, feature engineering (knee shape characterization) and model training could be performed in a fully automated way.

15.2.3 Predictive model

A custom python script was used to create an end-to-end machine learning pipeline in scikit-learn. Prior to model training, the dataset was split into a balanced training (n=180) and test (n=60) dataset, with equal proportions of responders and post-meniscectomy syndrome patients, equal patient sex ratios and equal BMI category ratios. A nested cross-validation approach was implemented, therefore the test set was alternated over 4 non-overlapping test sets by the outer cross-validation loop (see Figure 13.3 in section ”13.6.2. Evaluation Strategies”). The body mass index of every patient was calculated as their weight (kg) divided by their square height (m). Next, the body mass index was categorized into 4 categories: <25 kg/m^2, between 25 and 27,5 kg/m^2, between 27,5 and 30 kg/m^2, and >30 kg/m^2. The categorical data were one-hot (patient sex and presence of lateral meniscus lesion) and ordinal (BMI category) encoded and and all continuous clinical variables (patient age, weight and length) were normalized into z-scores (subtracted by their means and divided by their standard deviations). As there were no missing data, no imputation strategy was required. The registered shape information was fed into the pipeline as a large matrix where the rows are the specific cases and the columns the x,y and z coordinates of the surface points. Surface point coordinates of the distal femur were transformed to their principal component scores, based on the learned principal components in the training set. Likewise, the surface point coordinates of the proximal tibia were also transformed to their principal component scores. The number of modes of shape variation was optimized during the hyperparameter optimization. These principal component scores were then concatenated and an additional PCA transformation was applied to obtain the principal component scores of distal femur and proximal tibia combined (again the

principal components itself were only learned on the training set). Of note, as PCA was applied on the PC weights of individual shape models no influence of relative position of femur to tibia was introduced. A complete overview of the model pipeline is given in Figure 14.1. An inner loop of cross-validation was applied during the randomized search for hyperparameter optimization (500 iterations times 4 outer cross-validation folds times 4 inner cross-validation folds = 8000 iterations per assessed classifier), as part of the previously described nested cross-validation approach. Multiple classifier algorithms were trained and evaluated in leave-one-out cross-validation experiments on the training set. A complete overview of the assessed classifier algorithms, as well as the search space for hyperparameter optimization is given in Table 14.3. Model evaluation involves assessment of the receiver operating characteristic (ROC) curves and their area under the curve (AUC-ROC), sensitivity, specificity and prediction accuracy, as well as the respective learning curves to assess potential overfitting.

Table 15.1: Summary of the sampled distributions during the randomized search for hyperparameter optimization for every assessed classifier algorithm.

Classifier algorithm	Hyperparameter search space [min - max (sampled distribution)]
Logistic regression	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u), Logistic regression C: $10^{-4} - 10^{-1}$(l)
Linear SVM	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u), SVM C: $10^{-4} - 10^{-1}$(l)
Radial basis function kernel SVM	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u), SVM C: $10^{-2} - 10$(l), SVM gamma: $10^{-4} - 10^{-1}$(l)
Gaussian Process Classifier	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u)
Decision Tree Classifier	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u) Max tree depth: $2 - 5$ (u)
Random Forest Classifier	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u) n estimators: $500 - 400$ (u), max samples per tree: $0.5 - 0.9$ (u), max samples per leaf: $0.1 - 0.4$ (u)
Adaboost Classifier	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u) n estimators: $100 - 500$ (u), learning rate: $10^{-3} - 10^{-1}$(l)
Naive Bayes Classifier	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u)
LDA Classifier	PCA Femur n components: $4 - 26$ (u), PCA Tibia n components: $4 - 26$ (u)
QDA Classifier	PCA Femur n components: $4 - 26$ (u), PCA tibia n components: $4 - 26$ (u) regularization parameter: $0.2 - 0.5$ (u)

u: sampled from uniform distribution; l: sampled from a log-uniform or reciprocal distribution
PCA: principal component analysis; SVM: support vector machine; LDA: linear discriminant analysis
QDA: quadratic discriminant analysis

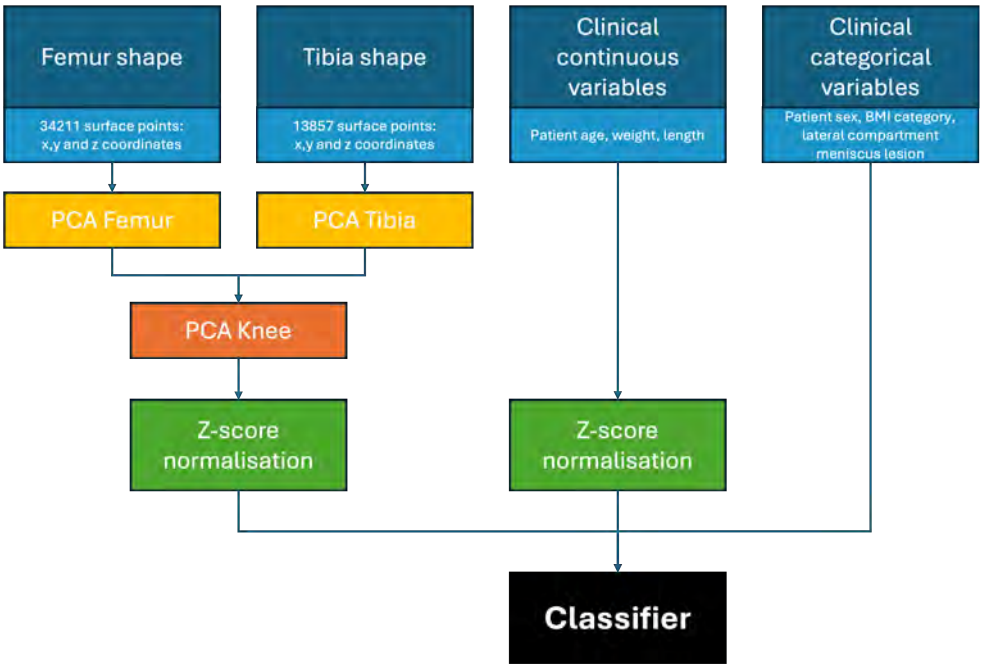


Figure 15.1: The applied scikit-learn pipeline streamlines the machine learning workflow by automating feature engineering and model training. By incorporating the feature engineering steps (e.g. consequent PCA transformations for statistical shape analysis) into one single pipeline, consistent data transformations and model evaluation are ensured across different experiments.

15.2.4 Model explainability

For future evolution towards a clinical decision support tool, the reasoning of the algorithm might be valuable to understand the algorithm's recommendation. This extra information is especially useful in the context of the shared decision making by the patient and the physician. A probability estimate as measure for the model's confidence, as well as a measure of each predictor feature's importance (model interpretation), give insights how to deal with the model's prediction. That way, low-confidence predictions can be overruled by the physician based on clinical expertise or by taking into account any exceptional patient characteristics not included by the current model. Therefore, the SHAP framework [29] was implemented post-hoc to calculate Shapley value approximations for each predictor feature. Similar to how the coefficients for every input predictor feature in linear or logistic regression models reflect their importance in simple regression problems/statistics, Shapley values assign predictor feature importance in a variety of more complex machine learning algorithms. In order to limit the complexity of our approach while still trying to avoid overfitting, no repeated cross-validation will be performed but the existing data splits of the outer cross-validation loop will be used to approximate the Shapley values. For every fold of the outer cross-validation, and for the features with the 15 largest Shapley value ranges (and thus influencing the final classification decision the most) the approximated

Shapley values will be visualized in a swarm plot. Finally, the three most important features to the classifier (see 14.2.3: Predictive model) were propagated backwards through the machine learning pipeline (to the original input format of bone surface point coordinates and clinical variables), allowing for the visualization of the most informative modes of bone shape variation.

15.3 Results

15.3.1 Clinical features

Both patient populations were first compared at group level to ensure comparable groups in terms of demographics and basic clinical factors such as patient sex, age, length and weight. No significant differences were observed, except for the distribution of BMI categories (Table 14.2).

Table 15.2: Descriptive statistics for patient age, sex, weight, length, BMI and BMI category for the R and MPMS group. Continuous variables were reported as means +/- standard deviation (SD), whereas categorical variables were reported as count (percentage).

		R group	**MPMS** group	p-value
Sex*	m	91 (75.8)	84 (70)	n.s.
	f	29 (24.2)	36 (30)	
Age [year]		50.6 (12.1)	53.1 (10.0)	n.s.
Weight [kg]		82.0 (15.1)	82.3 (14.3)	n.s.
Length [cm]		177.2 (8.6)	175.3 (9.3)	n.s.
BMI [kg/cm^2]		26.0 (3.8)	26.7 (3.3)	n.s.
BMI category†	<25	54 (45.0)	34 (28.33)	0.0496
	25-27.5	32 (26.7)	41 (34.2)	
	27.5-30	19 (15.8)	21 (17.5)	
	>30	15 (12.5)	24 (20)	

*: Fisher's Exact test, †: Pearson's chi-squared test
R: response to treatment group; MPMS: medial post-meniscectomy syndrome group
n.s.: not significant

15.3.2 Predictive model

Sensitivity and specificity to predict the MPMS onset, as well as prediction accuracy were summarized in Table 14.3. These metrics were calculated from the validation subsets in the inner cross-validation splits (4-fold), as none of the test data from the outer cross-validation loop could be used for final model selection, but only for final model evaluation. These were further aggregated over the outer cross-validation loop and reported as mean +/- standard deviation (SD). Eventual overfitting was visually evaluated on the learning curves, which are shown for the finally selected model in Figure 14.2. Overall, the logistic regression classifier performed most consistently over

the four outer cross-validation folds for all assessed metrics (sensitivity, specificity, accuracy and ROC-AUC), while showing no signs of overfitting. Therefore, in addition to its simplicity, this classifier was selected as final classifier and evaluated further.

Table 15.3: Predictive model results: mean and SD over the 4 folds of the outer cross-validation loop (and calculated in the inner 4-fold cross-validation).

Classifier algorithm	MPMS sens.	MPMS spec.	Acc.	ROC-AUC
Log regression	0.750 (0.019)	0.809 (0.027)	0.786 (0.014)	0.815 (0.017)
Linear SVM	0.736 (0.026)	0.818 (0.041)	0.785 (0.015)	0.820 (0.008)
RBF SVM*	0.744 (0.009)	0.793 (0.055)	0.774 (0.032)	0.825 (0.021)
GP Classifier	0.764 (0.033)	0.805 (0.026)	0.789 (0.022)	0.808 (0.017)
DT Classifier	0.592 (0.092)	0.691 (0.046)	0.660 (0.021)	0.680 (0.018)
RF Classifier	0.728 (0.049)	0.755 (0.055)	0.744 (0.044)	0.765 (0.044)
Adaboost Classifier*	0.697 (0.037)	0.758 (0.042)	0.736 (0.028)	0.775 (0.053)
NB Classifier	0.781 (0.021)	0.676 (0.033)	0.703 (0.032)	0.743 (0.037)
LDA	0.758 (0.019)	0.780 (0.045)	0.771 (0.023)	0.795 (0.017)
QDA*	0.731 (0.028)	0.756 (0.028)	0.747 (0.023)	0.785 (0.031)

MPMS: medial post-meniscectomy syndrome; sens.: sensitivity; spec.: specificity; acc.: accuracy;
ROC-AUC: receiver operating characteristic curve - area under the curve
Log regression: logistic regression; SVM: support vector machine; RBF: radial basis function kernel;
GP: Gaussian Process; DT: Decision Tree; RF: Random Forest; NB: Naive Bayes;
LDA: linear discriminant analysis; QDA: quadratic discriminant analyis
*learning curve suggests minor overfitting

To assess generalizability of prediction performance to unseen data by the model, results of the outer 4-fold cross-validation are summarized in Table 14.4. Additionally, the receiver operating characteristic (ROC) curves on the four test folds are plotted in Figure 14.3.

Table 15.4: Predictive model results in the 4 test folds of the outer cross-validation loop.

Test fold	MPMS sens.	MPMS spec.	Acc.	ROC-AUC
1	0.700	0.808	0.767	0.790
2	0.800	0.750	0.767	0.850
3	0.667	0.833	0.767	0.820
4	0.800	0.800	0.800	0.780

MPMS: medial post-meniscectomy syndrome; sens.: sensitivity; spec.: specificity; acc.: accuracy;
ROC-AUC: receiver operating characteristic curve - area under the curve

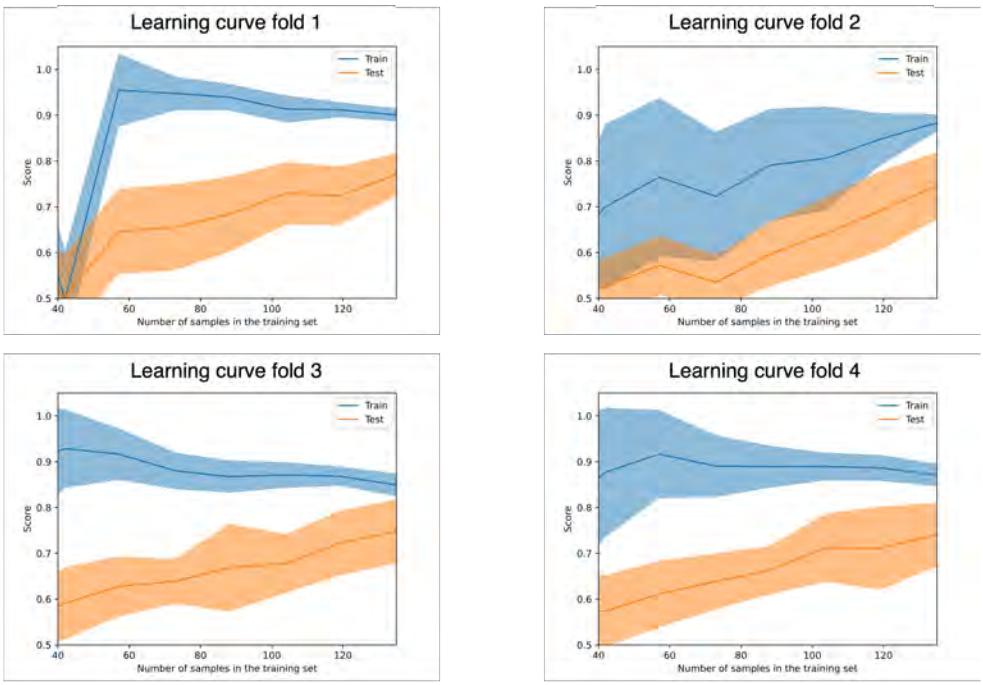


Figure 15.2: Learning curves for the logistic regression classifier (final model). Prediction accuracy (score) increased on the test dataset (orange curves) with an increasing size of the training dataset, while the performance on the training set cases (blue curves) decreased a bit. The rising trend of the test learning curves suggests potential to improve prediction accuracy by including more cases in the training dataset.

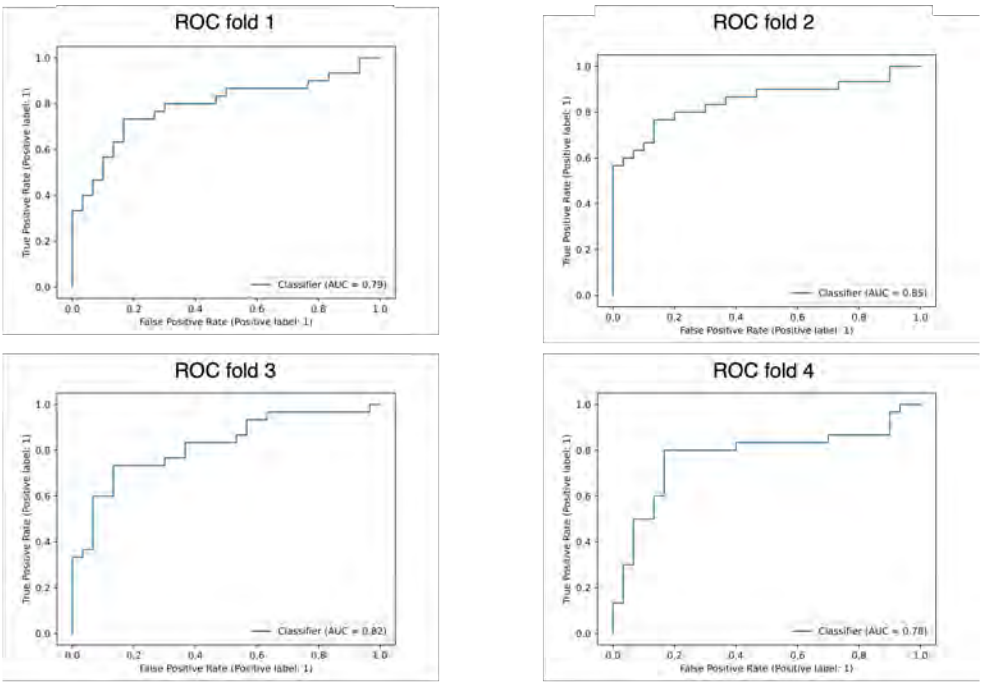


Figure 15.3: ROC curves for the logistic regression classifier on the four non-overlapping test folds (outer cross-validation loop).

15.3.3 Model explainability

The Shapley values in the four outer cross-validation folds were summarized in the swarm plots in Figure 14.4. The findings reveal that all input variables contribute to the final decision, with morphological features being the greatest impact assigned to by the model. Remark the different number of input features to the classifier, as the number of principal components for the femur and tibia were optimized during each hyperparameter optimization and hence dependent on the different validation sets. Over all outer cross-validation iterations, the principal component with index 0, 2 and 3 were attributed the largest influence to the final decision by the model. These principal components or modes of shape variation are visualized in Figures 14.5, 14.6 and 14.7. It comprises a smaller knee size (Figure 14.5), a wider mediolateral notch (femur) and intercondylar eminence (tibia) width (Figure 14.6) and a smaller medial femoral condyle size combined with some vague shape variations at the intercondylar tibial eminence (Figure 14.7) as risk factors for MPMS.

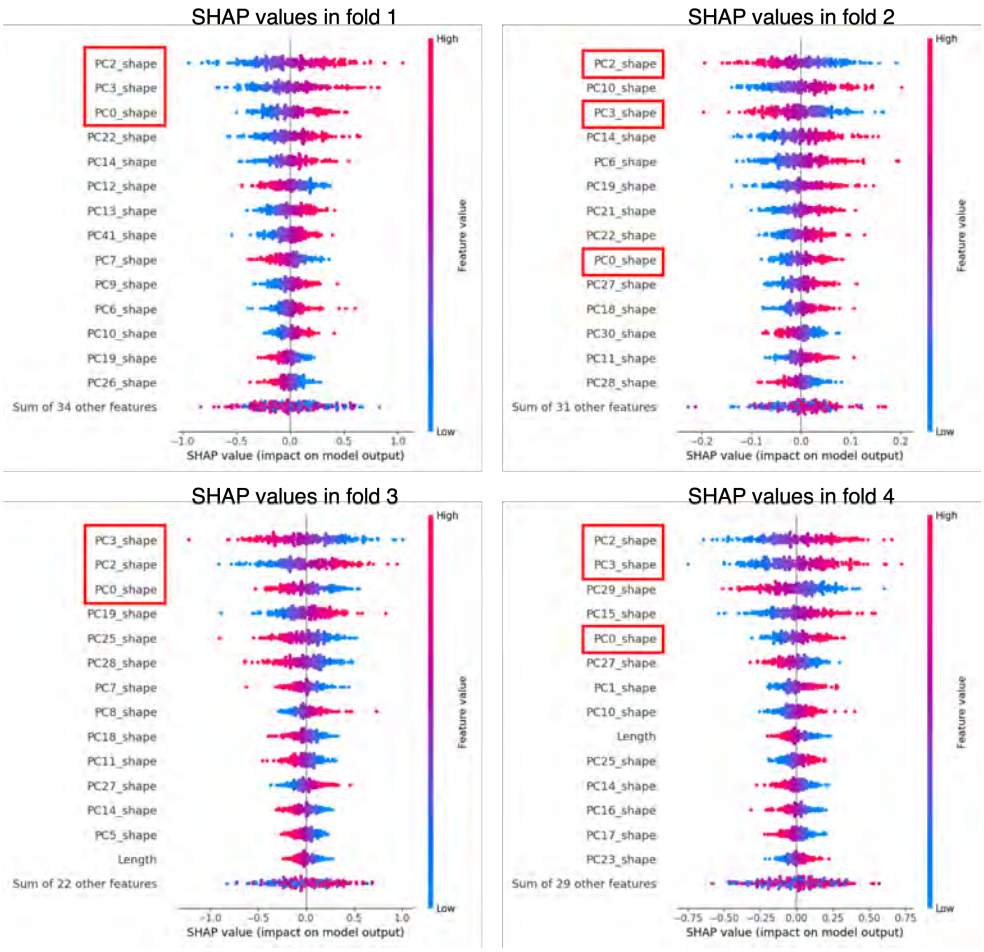


Figure 15.4: Shapley values for the 4-fold trained logistic regression models. The principal components with index 0,2 and 3 showed the largest range in Shapley values over the four folds of the outer cross-validation loop.

Figure 15.5: Visualization of the first principal component (index 0), representing the main mode of knee shape variation (= threedimensional size). The left column illustrates a shape deviating three standard deviations from the mean, aligned with the trend towards treatment response (R group, larger knee). The center column depicts the mean shape, while the right column shows a shape three standard deviations from the mean, associated with increased risk for MPMS (smaller knee).

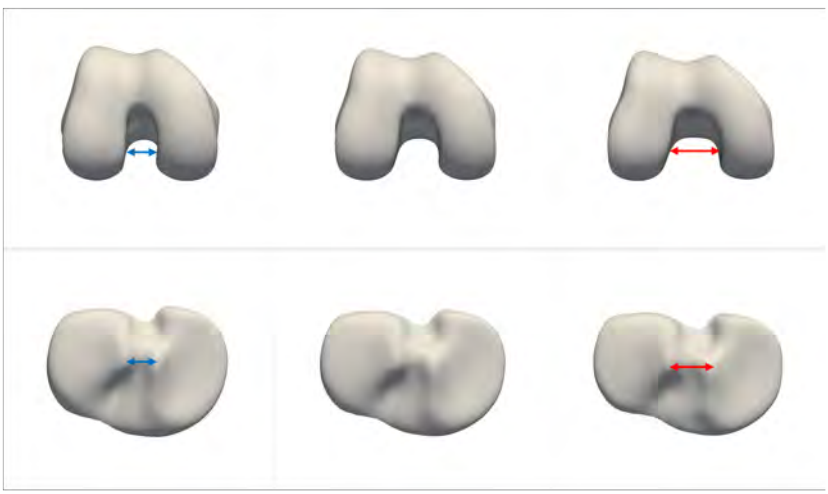


Figure 15.6: Visualization of the third principal component (index 2), representing a mode of knee shape variation (= notch width). The left column illustrates a shape deviating three standard deviations from the mean, aligned with the trend towards treatment response (R group, smaller intercondylar notch and tibial eminence). The center column depicts the mean shape, while the right column shows a shape three standard deviations from the mean, associated with increased risk for MPMS (mediolaterally wider intercondylar notch and tibial eminence)

Figure 15.7: Visualization of the fourth principal component (index 3), representing a mode of knee shape variation (= medial femoral condyle width). The left column illustrates a shape deviating three standard deviations from the mean, aligned with the trend towards treatment response (R group, smaller medial femoral condyle, anteroposterior deeper femoral notch). The center column depicts the mean shape, while the right column shows a shape three standard deviations from the mean, associated with increased risk for MPMS (wider medial femoral condyle, anterioposteriorly smaller intercondylar notch). At the level of the tibia this mode of shape variation is concentrated at the intercondylar region and slightly affects the relative position of the medial to the lateral tibial plateau in the axial plane.

15.4 Discussion

To the author's knowledge, the proposed method involves the first knee morphology aware predictive model for arthroscopic partial medial meniscectomy. Multiple models, each of them with its own advantages and disadvantages, were assessed in an extensive nested cross-validation experiment. A high sensitivity (74.2%), specificity (79,8%), accuracy (77.5%) and ROC-AUC (0.81) to predict the MPMS on previously unseen cases was observed. Additionally, learning curves suggest further improvement of the predictive performance by training the model on a larger dataset.

This study represents a pioneering integration of automated knee bone morphology analysis with conventional clinical data into a unified machine learning pipeline. The results are in line with our previous exploratory study on shape variations in the same dataset [27]. Again, the same modes of shape variations were found to be associated with the development of MPMS: a smaller knee size, a wider mediolateral femoral notch and a smaller medial femoral condyle.

The achieved discriminative power of our model outperformed the study of Lu et al. on graft failure and contralateral ACL injury prediction [30] in a dataset of 1497 patients. Unfortunately, no results on prediction sensitivity, specificity or accuracy were reported, and only the ROC-AUC was reported as being 0.67 for contralateral ACL injury and 0.70

for graft failure. Another study employing machine learning for predictive models related to knee arthroscopy attempted to predict opioid use following arthroscopy on a dataset of 381 patients [31]. An ensemble model of a support vector machine, random forest, extreme gradient boost and adaptive boosting classifier resulting in a ROC-AUC of 0.74. An imaging-based study by Zhao et al. successfully predicted the recurrence of hallux valgus based solely on radiographic measurements [32]. Their approach utilized features extracted from both pre-operative weightbearing and post-operative non-weightbearing images as inputs for a support vector machine classifier, achieving an accuracy of 75.6% and an ROC-AUC of 0.88. A systematic review on predictive models for clinical outcomes in hip arthroscopy highlighted considerable variability in model performance [33], with reported ROC-AUC values ranging from 0.57 [34] to 0.94 [35], depending on the predictor variables used (radiographic indices alone or combined with clinical data). In shoulder arthroplasty, machine learning models have already achieved over 90% accuracy in predicting 30-day post-operative complications [36, 37]. Notably, these studies benefited from a large dataset of 21,544 elective total shoulder arthroplasty cases (80% for training, 20% for testing), enabling the use of more complex artificial neural networks.

One notable strength of this study was the rigorous approach taken to ensure the completeness of patient records during the inclusion process. By systematically excluding records with incomplete data, the study eliminated the need for data imputation, thus avoiding potential sources of noise or bias associated with such methods [38]. Additionally, given the relatively small size of the dataset, the implementation of a nested cross-validation approach enabled the utilization of all available data throughout each stage of model development. This approach minimized the risk of introducing optimistic bias, which can occur when tuning model (hyper)parameters on the same dataset used for assessing model performance [39]. Particularly in smaller test datasets, the observed model performance can be affected by the inherent randomness involved in partitioning data into training and test subsets. By leveraging all non-overlapping test splits for model evaluation, this study reduced the likelihood of unintentionally selecting subsets of data that might be inherently more challenging or easier, ensuring each subject was used once in testing the model's performance [40].

A common challenge in this study, as with all machine learning approaches, is the need for high-quality, sufficiently large training datasets. Moreover, it is essential that the training data accurately reflect the patient population for which the model is intended. Ideally, this requires the dataset to show feature distributions that align closely with those of the target population, providing enough variability to support robust learning and generalizability. Although our dataset of 240 post-meniscectomy patients may be considered sufficiently large, the learning curves suggest that model performance could improve with additional training samples. Additionally, external validation is crucial to confirm the study's findings. These independent validations are essential to strengthen clinician confidence, supporting the model's integration into clinical practice [33, 41]. Next, the implemented calculation of Shapley values provides a popular approach for interpreting complex predictive models by attributing each predictor variable's contribution to the model's output [42, 43, 44, 45, 46]. However, Shapley values face limitations, particularly when predictor variables are correlated. As previously reported [27], strong correlations between knee shape and clinical data are present, e.g. linking knee size to patient length, or knee aspect ratio (mediolateral/anteroposterior)

to biological sex [47]. Additionally, patient weight and length are also strongly correlated. In those cases, Shapley values can misattribute importance across predictors, as the method lacks a mechanism for disentangling shared influence among correlated variables. This can lead to misleading interpretations, where correlated variables either receive inflated or diluted importance scores depending on the distribution of their interdependencies [48]. In contrast, causal inference models aim to model relationships that more explicitly account for causal pathways and conditional independence, often using assumptions or data-driven methods to control for confounding [49]. Causal machine learning is an emerging subdiscipline of machine learning, with still some challenges before translation to clinical practice [50].

Future research directions could focus on integrating biomechanical simulations (e.g. finite element analysis) to further interpret these data-driven predictor values. Given that many of the predictor features significantly influence knee biomechanics, such simulations would offer a physically grounded rationale for the observed data-driven outcomes. Moreover, reverse engineering approaches could potentially define patient-specific targets for weight management or adjustments in activity patterns. Additionally, incorporating kinematic and kinetic data from gait analysis as preoperative predictor variables may further strengthen the model's predictive capability [51]. Currently, the model relies on static analysis alone, despite evidence indicating the crucial role of the knee's soft tissue envelope for stability purposes—an aspect not yet represented in the existing model.

15.5 Conclusion

In conclusion, this study represents a promising initial step toward developing a clinical decision support system for the treatment of medial meniscus lesions. On average, the proposed machine learning model achieved an accuracy of 77.5% in the prediction of the medial post-meniscectomy syndrome. Following analysis of the Shapley values, following shape variations affected the predicted risk for medial post-meniscectomy the most: a smaller knee size, a wider femoral intercondylar notch, and a smaller medial femoral condyle. All of these shape variations are related to the medial compartment contact surface. A stronger evidence-based patient selection for arthroscopic partial medial meniscectomy has the potential to enhance surgical success rates while reducing the risk of medial post-meniscectomy syndrome and, consequently, delaying or preventing the early onset of medial knee osteoarthritis.

Bibliography

[1] Thorlund JB, Hare KB, Lohmander LS. Large increase in arthroscopic meniscus surgery in the middle-aged and older population in Denmark from 2000 to 2011. Acta orthopaedica. 2014 6;85:287-92.

[2] Beaufils P, Becker R, Kopf S, Matthieu O, Pujol N. The knee meniscus: management of traumatic tears and degenerative lesions. EFORT open reviews. 2017 5;2:195-203.

[3] Hall MJ, Schwartzman A, Zhang J, Liu X. Ambulatory Surgery Data From Hospitals and Ambulatory Surgery Centers: United States, 2010. National health statistics reports. 2017 2:1-15.

[4] Katano H, Koga H, Ozeki N, Otabe K, Mizuno M, Tomita M, et al. Trends in isolated meniscus repair and meniscectomy in Japan, 2011–2016. Journal of Orthopaedic Science. 2018 7;23:676-81.

[5] Jacquet C, Pujol N, Pauly V, Beaufils P, Ollivier M. Analysis of the trends in arthroscopic meniscectomy and meniscus repair procedures in France from 2005 to 2017. Orthopaedics & Traumatology: Surgery & Research. 2019 6;105:677-82.

[6] Beaufils P, Becker R, Kopf S, Englund M, Verdonk R, Ollivier M, et al. Surgical management of degenerative meniscus lesions: the 2016 ESSKA meniscus consensus. Knee Surgery, Sports Traumatology, Arthroscopy. 2017 2;25:335-46.

[7] Fairbank TJ. Knee joint changes after meniscectomy. The Journal of Bone and Joint Surgery British volume. 1948 11;30-B:664-70.

[8] Roemer FW, Kwoh CK, Hannon MJ, Hunter DJ, Eckstein F, Grago J, et al. Partial meniscectomy is associated with increased risk of incident radiographic osteoarthritis and worsening cartilage damage in the following year. European Radiology. 2017 1;27:404-13.

[9] Sihvonen R, Paavola M, Malmivaara A, Itälä A, Joukainen A, Kalske J, et al. Arthroscopic partial meniscectomy for a degenerative meniscus tear: A 5 year follow-up of the placebo-surgery controlled FIDELITY (Finnish Degenerative Meniscus Lesion Study) trial. British Journal of Sports Medicine. 2020.

[10] Englund M, Guermazi A, Lohmander SL. The Role of the Meniscus in Knee Osteoarthritis: a Cause or Consequence? Radiologic Clinics of North America. 2009 7;47:703-12.

[11] Paradowski PT, Lohmander LS, Englund M. Osteoarthritis of the knee after meniscal resection: long term radiographic evaluation of disease progression. Osteoarthritis and Cartilage. 2016 5;24:794-800.

[12] Pihl K, Ensor J, Peat G, Englund M, Lohmander S, Jørgensen U, et al. Wild goose chase – no predictable patient subgroups benefit from meniscal surgery: patient-reported outcomes of 641 patients 1 year after surgery. British Journal of Sports Medicine. 2020 1;54:13-22.

[13] Zaslav KR, Farr J, Alfred R, Alley RM, Dyle M, Gomoll AH, et al. Treatment of post-meniscectomy knee symptoms with medial meniscus replacement results in greater pain reduction and functional improvement than non-surgical care. Knee Surgery, Sports Traumatology, Arthroscopy. 2022 4;30:1325-35.

[14] Drobnič M, Ercin E, Gamelas J, Papacostas ET, Slynarski K, Zdanowicz U, et al. Treatment options for the symptomatic post-meniscectomy knee. Knee Surgery, Sports Traumatology, Arthroscopy. 2019 6;27:1817-24.

[15] Kohli S, Schwenck J, Barlow I. Failure rates and clinical outcomes of synthetic meniscal implants following partial meniscectomy: a systematic review. Knee surgery & related research. 2022 6;34:27.

[16] Wu KA, Therien AD, Kiwinda LV, Castillo CJ, Hendren S, Long JS, et al. Addressing meniscal deficiency part 2: An umbrella review of systematic reviews and meta-analyses on meniscal scaffold-based approaches. Journal of Experimental Orthopaedics. 2024 JUL;11(3).

[17] Hinterwimmer F, Lazic I, Langer S, Suren C, Charitou F, Hirschmann MT, et al. Prediction of complications and surgery duration in primary TKA with high accuracy using machine learning with arthroplasty-specific data. Knee Surgery, Sports Traumatology, Arthroscopy. 2023;31(4):1323-33.

[18] Willinger L, Lang JJ, Berthold D, Muench LN, Achtnich A, Forkel P, et al. Varus alignment aggravates tibiofemoral contact pressure rise after sequential medial meniscus resection. Knee Surgery, Sports Traumatology, Arthroscopy. 2020 4;28:1055-63.

[19] Bakke D, Besier T. Shape model constrained scaling improves repeatability of gait data. Journal of Biomechanics. 2020;107:109838.

[20] Willems M, Killen BA, Di Raimondo G, Van Dijck C, Havashinezhadian S, Turcot K, et al. Population-based in silico modeling of anatomical shape variation of the knee and its impact on joint loading in knee osteoarthritis. Journal of Orthopaedic Research. 2024;42(11):2473-84.

[21] Jiang J, Liu Z, Wang X, Xia Y, Wu M. Increased Posterior Tibial Slope and Meniscal Slope Could Be Risk Factors for Meniscal Injuries: A Systematic Review. Arthroscopy. 2022 7;38:2331-41.

[22] Oevelen AV, den Borre IV, Duquesne K, Pizurica A, Victor J, Nauwelaers N, et al. Wear patterns in knee OA correlate with native limb geometry. Frontiers in bioengineering and biotechnology. 2022;10:1042441.

[23] Paz A, García J, Korhonen RK, Mononen ME. Towards a Transferable Modeling Method of the Knee to Distinguish Between Future Healthy Joints from Osteoarthritic Joints: Data from the Osteoarthritis Initiative. Annals of Biomedical Engineering. 2023;51(10):2192-203.

[24] Abbadessa A, Nuñez Bernal P, Buttitta G, Ronca A, D'Amora U, Zihlmann C, et al. Biofunctionalization of 3D printed collagen with bevacizumab-loaded microparticles targeting pathological angiogenesis. Journal of Controlled Release. 2023;360:747-58.

[25] Levato R, Jungst T, Scheuring RG, Blunk T, Groll J, Malda J. From Shape to Function: The Next Step in Bioprinting. Advanced Materials. 2020;32(12):1906423.

[26] Ghisa C, Zaslav KR. Current state of off the shelf scaffolds and implants for meniscal replacement. Journal of Cartilage & Joint Preservation. 2022 3;2:100040.

[27] Grammens J, Haver AV, Danckaers F, Vuylsteke K, Sijbers J, Mahluf L, et al. Three-dimensional bone morphology is a risk factor for medial postmeniscectomy syndrome: A retrospective cohort study. Journal of Experimental Orthopaedics. 2024 7;11.

[28] Danckaers F, Huysmans T, Lacko D, Ledda A, Verwulgent S, Dongen SV, et al. Correspondence Preserving Elastic Surface Registration with Shape Model Prior. In: 2014 22nd International Conference on Pattern Recognition. IEEE; 2014. p. 2143-8.

[29] Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017. .

[30] Lu Y, Till SE, Labott JR, Reinholz AK, Hevesi M, Krych AJ, et al. Graft Failure and Contralateral ACL Injuries After Primary ACL Reconstruction: An Analysis of Risk Factors Using Interpretable Machine Learning. Orthop J Sports Med. 2024 Oct;12(10):23259671241282316.

[31] Lu Y, Forlenza E, Wilbur RR, Lavoie-Gagne O, Fu MC, Yanke AB, et al. Machine-learning model successfully predicts patients at risk for prolonged postoperative opioid use following elective knee arthroscopy. Knee Surgery, Sports Traumatology, Arthroscopy. 2022;30(3):762-72.

[32] Zhao R, Wang G, Li F, Wang J, Zhang Y, Li D, et al. Developing Machine Learning-Based Predictive Models for Hallux Valgus Recurrence Based on Measurements From Radiographs. Foot Ankle Int. 2024 Sep;45(9):1000-8.

[33] Mehta A, El-Najjar D, Howell H, Gupta P, Arciero E, Marigi EM, et al. Artificial Intelligence Models Are Limited in Predicting Clinical Outcomes Following Hip Arthroscopy: A Systematic Review. JBJS Reviews. 2024;12(8).

[34] Ramkumar PN, Karnuta JM, Haeberle HS, Sullivan SW, Nawabi DH, Ranawat AS, et al. Radiographic Indices Are Not Predictive of Clinical Outcomes Among 1735 Patients Indicated for Hip Arthroscopic Surgery: A Machine Learning Analysis. The American Journal of Sports Medicine. 2020;48(12):2910-8. PMID: 32924530.

[35] Kunze KN, Polce EM, Rasio J, Nho SJ. Machine Learning Algorithms Predict Clinically Significant Improvements in Satisfaction After Hip Arthroscopy. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2021;37(4):1143-51.

[36] Lopez CD, Constant M, Anderson MJJ, Confino JE, Heffernan JT, Jobin CM. Using machine learning methods to predict nonhome discharge after elective total shoulder arthroplasty. JSES International. 2021;5(4):692-8.

[37] Lopez CD, Constant M, Anderson MJ, Confino JE, Lanham NS, Jobin CM. Using machine learning methods to predict prolonged operative time in elective total shoulder arthroplasty. Seminars in Arthroplasty: JSES. 2022;32(3):452-61.

[38] Thuraisingam S, Chondros P, Dowsey MM, Spelman T, Garies S, Choong PF, et al. Assessing the suitability of general practice electronic health records for clinical prediction model development: a data quality assessment. BMC Medical Informatics and Decision Making. 2021;21(1):297.

[39] Wilimitis D, Walsh CG. Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial. JMIR AI. 2023 Dec;2:e49023.

[40] Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006;7(1):91.

[41] Sanghvi PA, Shah AK, Hecht CJ, Karimi AH, Kamath AF. Optimal inputs for machine learning models in predicting total joint arthroplasty outcomes: a systematic review. Eur J Orthop Surg Traumatol. 2024 Dec;34(8):3809-25.

[42] Lundberg SM, Lee S. A unified approach to interpreting model predictions. CoRR. 2017;abs/1705.07874. Available from: http://arxiv.org/abs/1705.07874.

[43] Wang H, Doumard E, Soulé-Dupuy C, Kémoun P, Aligon J, Monsarrat P. Explanations as a New Metric for Feature Selection: A Systematic Approach. IEEE Journal of Biomedical and Health Informatics. 2023;27:4131-42.

[44] Fan Y, Mao H, Li Q. A Model-Agnostic Feature Attribution Approach to Magnetoencephalography Predictions Based on Shapley Value. IEEE Journal of Biomedical and Health Informatics. 2023;27:2524-35.

[45] Zhuang J, Huang H, Jiang SF, Liang J, Liu Y, Yu X. A generalizable and interpretable model for mortality risk stratification of sepsis patients in intensive care unit. BMC Medical Informatics and Decision Making. 2023;23.

[46] min Park H, Yun I, Kim M, Nguyen KT, Messem AV, Neve WD. Interpretable rotator cuff tear diagnosis using MRI slides with CAMscore and SHAP. In: Medical Imaging; 2024. .

[47] Audenaert EA, Pattyn C, Steenackers G, De Roeck J, Vandermeulen D, Claes P. Statistical Shape Modeling of Skeletal Anatomy for Sex Discrimination: Their Training Size, Sexual Dimorphism, and Asymmetry. Front Bioeng Biotechnol. 2019;7:302.

[48] Aas K, Jullum M, Løland A. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. Artificial Intelligence. 2021;298:103502.

[49] Lundberg S. Be Careful When Interpreting Predictive Models in Search of Causal Insights. Medium.com; 2021. Available from: https://towardsdatascience.com/be-careful-when-interpreting-predictive-models-in-search-of-causal-insights-e68626e664b6.

[50] Feuerriegel S, Frauen D, Melnychuk V, Schweisthal J, Hess K, Curth A, et al. Causal machine learning for predicting treatment outcomes. Nature Medicine. 2024;30(4):958-68.

[51] Liew BXW, Pfisterer F, Rügamer D, Zhai X. Strategies to optimise machine learning classification performance when using biomechanical features. Journal of Biomechanics. 2024;165:111998.

Part IV

Conclusion and future perspectives

Chapter 16
Conclusions

16.1 Technological findings

16.1.1 Automated image segmentation

During the collection of multi-center pre-operative MRI imaging of meniscus patients, image quality presented the first major challenge of this research. This was addressed through an exploration of currently available MRI knee scan protocols, and extended towards the main imaging modalities for diagnostic purposes in the field of orthopaedics (CT and MRI). Both imaging modalities are commonly used to create 3D models for any anatomy of interest. As previously discussed, the accuracy of machine learning outputs in any application is fundamentally limited by the quality (e.g. resolution) of its input data. Achieving optimal image quality in MRI requires careful balancing of parameters affecting image field of view and resolution, signal-to-noise ratio, and scan time. Conventionally used clinical 2D MRI protocols, which typically involve multiple sequences acquired in different planes (e.g. sagittal, axial,...) with thicker slices, were examined in conjunction with recent advancements in 3D MRI protocols for the assessment of knee morphology and thinner anatomical structures such as cartilage layers. Near-isotropic 3D MRI protocols, which provide high-resolution images with minimal slice thickness, were identified as particularly well-suited for the automated segmentation of knee bone and cartilage. A second challenge addressed in this study was the labor-intensive and time-consuming nature of manual image segmentation, a bottleneck for scaling up knee morphology analysis towards routine clinical applications. Since the introduction of the U-Net architecture in 2015, which resulted in a revolution in the field image segmentation, a wide range of neural network architectures has been proposed to automate segmentation tasks across diverse datasets. Four state-of-the-art neural networks were carefully selected for comparison in the context of MRI knee segmentation.

16.1.2 Automated anatomical landmarking

Based on 3D models derived from pre-operative MRI scans of meniscus patients, a tool for automated landmarking was developed and validated. First, the manual method for

landmark-based morphometric measurements was validated, and consequently the difference between the automated method and the mean of three observers as consensus was assessed. The intermethod variability was in the same range as the interobserver variability and 71% and 95% of all automated morphometric measurements measurements were within 1 mm and 2 mm agreement with the manual consensus measurements. This landmarking tool has several potential applications, including donor-recipient matching for meniscal allograft transplantation (MAT) and morphometric analyses to select optimal component sizes in total knee arthroplasty (TKA).

16.1.3 Automated statistical shape analysis

Following the establishment of anatomical correspondences across 3D bone surfaces, statistical shape models can be constructed, allowing quantification of shape variations without the need for predefined landmarks. This data-driven approach enables the extraction of principal components (or main modes of shape variation) within a dataset of 3D bone surfaces, where each shape instance can be represented by a vector of principal component weights.

16.1.4 Automated prediction of response to treatment versus failure

Elementary concepts of machine learning were explored and summarized, to serve as background information for the final knee morphology-aware predictive model for arthroscopic partial medial meniscectomy outcome (= clinical finding, cf. next section).

16.2 Clinical findings

16.2.1 Clinical applications of 3D image analysis

Four state-of-the-art neural networks for image segmentation were compared, both from a computer science point of view using conventional segmentation metrics such as the Dice score, and from an anatomical point of view by means of statistical parametric mapping, in order to assess the anatomical location of erronous trends and maximal errors in the 3D model domain. This comparative analysis did not reveal remarkable insights with respect to segmentation accuracy, thereby shifting the focus towards practical implemenation and ease of use when choosing an automated MRI knee segmentation algorithm for bone and cartilage. Following 3D surface registration, several clinical applications can be considered, going from landmark-based measurements such as the femoral anteversion, to the reconstruction of missing anatomies.

16.2.2 3D knee morphology as risk factor for symptomatic knee degeneration in the meniscus deficient knee

A pilot study was performed to investigate an empirical observation from clinics: a subpopulation of medial compartment degeneration patients at a young age, presenting with none of the known risk factors, but a rather narrow medial femoral condyle on cross-sectional views of the MRI scan. In this pilot study, both a landmark-based method and statistical parametric mapping were applied to characterize the small medial femoral condyle morphotype as follows: an effectively smaller medial femoral condyle and medial tibial plateau; a wider lateral femoral condyle and a wider distal femur on a smaller tibial plateau. This finding was further backed by a next study, comparing the characteristics of donor (functional meniscus) and acceptor (dysfunctional meniscus) meniscal allografts in MAT. A discrepancy in terms of size was found for the medial meniscus allografts: meniscus deficient patients had smaller medial menisci than the general donor population. This again associates a smaller medial tibiofemoral compartment to early meniscal failure in neutrally aligned knees (cf. indications for MAT). Next, a large retrospective cohort study of 240 post-meniscectomy patients was conducted in the context of the MEFISTO project, with an equal proportion of responders to arthroscopic partial medial meniscectomy and patients who developed the medial post-meniscectomy syndrome. The two groups were comparable in terms of patient sex, weight and length, but showed significant morphological differences. Following statistical shape analysis, the first three principal components (or main modes of shape variation) were significantly different between the responders and the MPMS patients. A smaller three-dimensional knee size, a wider femoral notch (and intercondylar eminence with for the tibia) and a smaller medial femoral condyle were significant characteristics of the MPMS group patients.

16.2.3 Optimal patient selection for arthroscopic partial medial meniscectomy

To predict the onset of MPMS on a patient-specific level, our morphotype-based predictive model was able to discriminate between responders and MPMS patients with a mean accuracy of 77,5% , a mean sensitivity for MPMS of 74.2% and a mean specificity for MPMS of 79.8% in a nested cross-validation experiment. To conclude, this first attempt to predict the outcome of APMM already paves the way for a morphotype-based clinical decision support tool for APMM, aiding the clinician by means of a data-driven patient selection, potentially resulting in higher surgery success rates and less failures caused by meniscal deficiency as observed in the post-meniscectomy syndrome.

16.3 Closing remark

To quote Ronald S. Weinstein (1938-2021), an academic pathologist:

> ***"A fool with a tool is still a fool."***

This statement can be interpreted in two ways:

1. "No tool can surpass the intelligence of the person using it."
2. "Using the wrong tool, or using the tool wrongly, offers no advantage."

While we are still some steps away from achieving artificial general intelligence (AGI), the relevance of the first interpretation is already open to debate. The second interpretation, however, is more pertinent than ever. Now, after the initial excitement for artificial intelligence, and particularly generative deep learning models such as GPT, some of the first related challenges have begun to emerge. Generative AI models as an example can sometimes be unreliable, prone to hallucination, or exhibit biases due to the data they were trained on. One should realize that AI and machine learning algorithms can only be applied effectively within the scope for which they were trained. If the target population has a significantly skewed distribution of a latent variable compared to the training dataset, the model's performance could degrade to little more than random guessing, based on basic probability theory. Therefore, it is essential to know and to keep in mind the final goal of the applied algorithm at all distinct phases of the development: from data collection and preprocessing to error analysis and final implementation into (clinical) practice. Some technical background knowledge of the end user might help in early detection of the model's limitations, and, more importantly, how to overcome them in the most efficient way. This PhD dissertation aims to be a first step into the implementation of data-driven predictive algorithms into orthopedic clinical practice. Just as medicine continuously evolves, both in terms of diagnosis and treatment options, these clinical decision support systems should evolve too, by adding more (recent) data to the training dataset to keep up with the current medical advances. Lifelong learning is not solely essential for healthcare professionals; it is equally vital for these data-driven systems, which must adapt and evolve alongside the ever-changing landscape of clinical practice.

Chapter 17

Future perspectives

17.1 Technological perspectives

The field of imaging is undergoing rapid advancements that extend beyond improving image quality, such as signal-to-noise ratio and resolution, in established modalities like CT and MRI. Innovations are emerging related to imaging hardware, acquisition protocols, software for image processing, and even patient positioning boundary conditions. One promising trend is low-field magnetic resonance imaging (MRI), which could make MRI more accessible in resource-limited settings [1]. With lower hardware demands, such as smaller and more affordable magnetic coils, and reduced requirements for superconducting conditions, low-field MRI may lower costs and expand availability. This approach not only offers logistical benefits but may also mitigate artifacts from metallic implants [2]. While low-field MRI is unlikely to directly improve the performance of imaging-based predictive models, the attention to lower magnetic fields in MRI is an important driver for the transition of innovations that were originally developed for high-field MRI towards regular clinical MRI scanners [3].

Another key area of innovation is the development of weight-bearing MRI protocols for the lower limb. Similar to recent advancements in computed tomography, weight-bearing MRI holds potential for assessing cartilage, menisci, and ligaments under load, thereby offering insight into the mechanical properties and structural response of knee tissues under compression [4]. Though still a simplification of real-world dynamic loading, weight-bearing MRI provides valuable data on joint behavior in a stance-like condition [5]. As both upright scanning in a standing position, as well as simulating load in supine position (by applying a predefined ground reaction force on the feet) are both challenging in MRI, weight-bearing CT is a highly valuable alternative for orthopaedic applications like weight-bearing lower limb alignment analysis under weight-bearing conditions [6].

Next, quantitative MRI (qMRI) holds significant promise as an advanced imaging modality for knee cartilage assessment, providing detailed insights into tissue composition and microstructural integrity that go beyond conventional MRI. By measuring specific biomarkers such as T1ρ, T2, T2*, and T1 relaxation times, qMRI enables the quantification of cartilage health, subchondral bone quality, and meniscal integrity—essential factors in understanding osteoarthritis (OA) progression and other

degenerative joint diseases [7, 8]. This quantitative approach allows for early detection of biochemical changes in cartilage and other knee joint tissues before morphological deterioration becomes visible, thus facilitating earlier diagnosis and intervention. As qMRI techniques continue to improve in speed and accuracy, including developments in deep learning [9] to enhance image reconstruction and interpretation, their clinical relevance is expanding, with applications in both patient stratification for treatment and monitoring therapy efficacy.

Phase-contrast X-ray imaging and phase-contrast computed tomography are other promising image modalities with potential future use cases in orthopaedics [10]. It makes use of the phase shift distribution, caused by X-ray refraction, and is therefore superior in weakly absorbing tissues, enabling microscopic images with spatial resolutions up to 0.2 nm [11]. It would therefore offer a non-invasive alternative to histological examination.

In the subsequent stage of image processing, deep learning is leading transformative changes. Deep learning algorithms have demonstrated remarkable efficacy in image reconstruction, allowing for reduced scan times while maintaining resolution and signal-to-noise ratios or superresolution [12]. Generative AI models further supports image interpretation by translating intricate visual details into interpretable formats, such as generating detailed text descriptions [13, 14] or flagging regions of clinical interest within images.

The domain of statistical shape analysis is also evolving with advancements in geometric deep learning, which leverages graph theory to extract anatomically meaningful features from complex shapes or point clouds [15, 16]. Geometric deep learning facilitates anatomical correspondence mapping on 3D surfaces and enables non-linear statistical shape analyisis, fostering a more comprehensive understanding of anatomical variations [17].

17.2 Clinical perspectives

Advancements in surgical techniques and breakthroughs in implant materials will continue to drive innovation in meniscus substitution therapy. Recently, the use of tendon autograft was investigated as an alternative for the scarce meniscal allografts. Promising results were already obtained in a rabbit model, showing chrondroprotective effects compared to the control group [18]. They accommodate for a wide range in size [19], thereby addressing the observed discrepancy in terms of meniscal size between acceptor demand and donor supply in MAT [20]. Surgical technique is also of crucial importance when replacing the non-functional native meniscus by an artificial meniscus implant, as it ensures the fixation of the implant to the tibial bone [21]. Additionally, research on 3D bioprinting and tissue engineering continues to fine-tune implant and scaffold materials, allowing for biomimetic implant designs aiming to restore the native joint homeostasis [22, 23, 24].

Healthcare applications are likely to see exponential growth in predictive and prognostic models as data availability increases, both in terms of feature count and patient volume. This expansion is anticipated to yield models with enhanced accuracy,

potentially evolving into clinical decision support systems (CDSS). These tools can provide clinicians with data-driven treatment recommendations, improving decision-making for individual patients [25, 26]. However, it will remain the clinician's responsibility to interpret model outputs, understanding the limits of predictive accuracy within the specific context of the model's training data and usage parameters. As healthcare systems integrate these advancements, they must be implemented with a strong emphasis on clinical applicability, safety, and interpretability, ensuring that predictive models support rather than replace the clinician's expertise.

A next step towards clinical implementation of the proposed methodology includes external validation by means of a large, prospective study. Extensive automation of image segmentation and statistical shape analysis makes the proposed workflow already scalable to massive datasets. The addition of other treatment options in a longitudinal follow-up study could even enable to extend the model from simply predicting the medial post-meniscectomy syndrome to a true clinical decision support system, directly suggesting the treatment option with the highest patient-specific chance of a successful outcome.

Bibliography

[1] Ayde R, Vornehm M, Zhao Y, Knoll F, Wu EX, Sarracanie M. MRI at low field: A review of software solutions for improving SNR. NMR Biomed. 2024 Oct:e5268.

[2] Luitjens J, Ziegeler K, Yoon D, Gassert F, Bhattacharjee R, Manatrakul R, et al. Improved metal suppression using new generation low-field MRI: a biophantom feasibility study. Skeletal Radiol. 2024 Oct.

[3] De Moura HL, Monga A, Zhang X, Zibetti MVW, Keerthivasan MB, Regatte RR. Feasibility of 3D MRI fingerprinting for rapid knee cartilage T1, T2, and T1ρ mapping at 0.55T: Comparison with 3T. NMR in Biomedicine. 2024;n/a(n/a):e5250.

[4] Bruno F, Barile A, Arrigoni F, Laporta A, Russo A, Carotti M, et al. Weight-bearing MRI of the knee: a review of advantages and limits. Acta Biomed. 2018 Jan;89(1-S):78-88.

[5] Severyns M, Zot F, Harika-Germaneau G, Germaneau A, Herpe G, Naudin M, et al. Extrusion and meniscal mobility evaluation in case of ramp lesion injury: a biomechanical feasibility study by 7T magnetic resonance imaging and digital volume correlation. Front Bioeng Biotechnol. 2023;11:1289290.

[6] Turmezei TD, B Low S, Rupret S, Treece GM, Gee AH, MacKay JW, et al. Quantitative Three-dimensional Assessment of Knee Joint Space Width from Weight-bearing CT. Radiology. 2021 Jun;299(3):649-59.

[7] Zhu H, Miller EY, Lee W, Wilson RL, Neu CP. In vivo human knee varus-valgus loading apparatus for analysis of MRI-based intratissue strain and relaxometry. J Biomech. 2024 Jun;171:112171.

[8] Ramsdell JC, Beynnon BD, Borah AS, Gardner-Morse MG, Zhang J, Krug MI, et al. Tibial and femoral articular cartilage exhibit opposite outcomes for T1ρand T2*

relaxation times in response to acute compressive loading in healthy knees. J Biomech. 2024 May;169:112133.

[9] Bian W, Jang A, Liu F. Improving quantitative MRI using self-supervised deep learning with model reinforcement: Demonstration for rapid T1 mapping. Magn Reson Med. 2024 Jul;92(1):98-111.

[10] Yang X, Shi L, Li A, Gao F, Sun W, Li Z. Phase-contrast imaging with synchrotron hard X-ray reveals the effect of icariin on bone tissue morphology and microstructure in rabbits with early glucocorticoid-induced osteonecrosis of the femoral head. Front Cell Dev Biol. 2023;11:1155532.

[11] Sun W, Zhang Y, Gao F, Li Z, Li G, Pan L. Phase-contrast imaging with synchrotron hard X-ray of micro lesions of the cartilage of the femoral head in rabbits. Int J Clin Exp Med. 2015;8(11):20086-91.

[12] Chaudhari AS, Fang Z, Kogan F, Wood J, Stevens KJ, Gibbons EK, et al. Super-resolution musculoskeletal MRI using deep learning. Magn Reson Med. 2018 Nov;80(5):2139-54.

[13] Dehghani N, Saliba-Colombani V, Chick A, Heng M, Operto G, Fillard P. Efficient labeling of french mammogram reports with MammoBERT. Sci Rep. 2024 Oct;14(1):24842.

[14] Santomartino SM, Zech JR, Hall K, Jeudy J, Parekh V, Yi PH. Evaluating the Performance and Bias of Natural Language Processing Tools in Labeling Chest Radiograph Reports. Radiology. 2024 Oct;313(1):e232746.

[15] Croquet B, Matthews H, Mertens J, Fan Y, Nauwelaers N, Mahdi S, et al. Automated landmarking for palatal shape analysis using geometric deep learning. Orthod Craniofac Res. 2021 Dec;24 Suppl 2(Suppl 2):144-52.

[16] Adams J, Elhabian S. Can point cloud networks learn statistical shape models of anatomies? Med Image Comput Comput Assist Interv. 2023 Oct;14220:486-96.

[17] Nauwelaers N, Matthews H, Fan Y, Croquet B, Hoskens H, Mahdi S, et al. Exploring palatal and dental shape variation with 3D shape analysis and geometric deep learning. Orthod Craniofac Res. 2021 Dec;24 Suppl 2(Suppl 2):134-43.

[18] Li C, Hu X, Meng Q, Zhang X, Zhu J, Dai L, et al. The potential of using semitendinosus tendon as autograft in rabbit meniscus reconstruction. Sci Rep. 2017 Aug;7(1):7033.

[19] Kim Y, Karl E, Ishijima M, Guy S, Jacquet C, Ollivier M. The Potential of Tendon Autograft as Meniscus substitution: Current concepts. J ISAKOS. 2024 Oct:100353.

[20] Tabbaa SM, Pace JL, Frank RM, Grammens J, Verdonk P. Meniscus Size Differs Between Patient and Donor Populations for Meniscus Allograft Transplantation. Arthrosc Sports Med Rehabil. 2023 Jun;5(3):e569-76.

[21] van Minnen BS, Heesterbeek PJC, Defoort KC, Emans PJ, van Arkel ERA, Struik T, et al. High failure rate but promising clinical performance after implantation of a flexible medial meniscus prosthesis at 1-year follow-up. Knee Surg Sports Traumatol Arthrosc. 2024 Sep.

[22] Kluyskens L, Debieux P, Wong KL, Krych AJ, Saris DBF. Biomaterials for meniscus and cartilage in knee surgery: state of the art. Journal of ISAKOS. 2022 2024/11/12;7(2):67-77.

[23] Klarmann GJ, Gaston J, Ho VB. A review of strategies for development of tissue engineered meniscal implants. Biomater Biosyst. 2021 Dec;4:100026.

[24] DePhillipo NN, LaPrade RF, Zaffagnini S, Mouton C, Seil R, Beaufils P. The future of meniscus science: international expert consensus. J Exp Orthop. 2021 Mar;8(1):24.

[25] Ningrum DNA, Kung WM, Tzeng IS, Yuan SP, Wu CC, Huang CY, et al. A Deep Learning Model to Predict Knee Osteoarthritis Based on Nonimage Longitudinal Medical Record. J Multidiscip Healthc. 2021;14:2477-85.

[26] Granviken F, Meisingset I, Bach K, Bones AF, Simpson MR, Hill JC, et al. Personalised decision support in the management of patients with musculoskeletal pain in primary physiotherapy care: a cluster randomised controlled trial (the SupportPrim project). Pain. 2024 Oct.

List of Figures

1.1 Anatomical reference planes 4

1.2 The patella as lever arm of the M. Quadriceps 5

1.3 Anatomy of the distal femur 6

1.4 J-curve of the knee in a sagittal cross-section 7

1.5 Anatomy of the proximal tibia and fibula 7

1.6 Human synovial joint 9

1.7 Knee ligaments and menisci 10

1.8 Anatomy of the medial and lateral meniscus 11

1.9 Femoral rollback and minor internal rotation during knee flexion 13

1.10 Coronal alignment of the knee 14

1.11 Hip-knee-angle (HKA), femoral mechanical angle (FMA) and tibial mechanical angle (TMA) 15

2.1 Arthroscopy image of a degenerative medial meniscus 20

2.2 Meniscus tear patterns 21

2.3 Functional meniscus deficiency consequences 23

2.4 Schematic overview of the mechanical risk factors hypothesized to drive knee degeneration 28

4.1 Simplified illustration of X-ray attenuation. 40

4.2 CT raw data acquisition and image reconstruction. 40

4.3 Application of CT-guided surgical planning: revision of a total knee prosthesis 41

4.4 Schematic overview of the basic MRI principles 43

4.5 Comparison of 2D and 3D MRI images. 45

4.6 Trade-offs to consider during MRI scan protocol design. 46

5.1 Manual image segmentation: screenshot from 3D Slicer (start of segmentation process) 51

5.2 Manual image segmentation: screenshot from 3D Slicer (end of segmentation process) 52

5.3 Overlay of the voxel grid of the scan volume containing the 3D anatomy of interest 54

5.4 Detail of a triangular 3D mesh structure of distal femur (upper mesh) and proximal tibia (lower mesh) 54

6.1 Schematic comparison between a biological neuron and a perceptron node. 58

6.2 Illustration of a neural network with two hidden layers 59

6.3 Schematic overview of the U-Net architecture. 60

6.4 Schematic overview of the automated configuration of the nnU-Net. . . . 63

6.5 Swin-UnetR model architecture. 65

6.6 Schematic overview of the MedNeXt architecture. 67

6.7 Schematic overview of the SegMamba model architecture. 69

6.8 Illustration of gated spatial convolutions and tri-orientated Mamba blocks. 69

7.1 Automated landmarking workflow overview. 76

7.2 Landmark visualization on the template bone and cartilage shapes. 78

7.3 Morphometric measurements visualized on the template bone and cartilage shapes. 80

7.4 Intra-observer, inter-observer and inter-method landmark position differences (box-and-whisker diagrams). 83

7.5 Intra-observer, inter-observer and inter-method measurement differences (box-and-whisker diagrams). 85

7.6 Success detection rates in % per landmark within predefined tolerance [mm]. 87

7.7 Success measurement rates in % per morphometric measurement within predefined tolerance [mm]. 89

7.8 Example of potential discrepancy between landmark position error and corresponding measurements error. 91

8.1 Visualisation of a subset of anatomical correspondences between source and target mesh. 97

8.2 Schematic overview of the iterative process employing shape model prior modulated registration. 100

9.1 Schematic overview of how machine learning relates to artificial intelligence (AI) and deep learning 110

9.2 Comparison between grid search and a randomized search strategy for hyperparameter optimization. 114

9.3 Schematic overview nested cross-validation with 4 outer and 4 inner folds. 118

10.1 Learning rate scheme 125

10.2 3D mapping of mean absolute segmentation errors for femoral bone. . . . 128

10.3 3D mapping of maximal absolute segmentation errors for femoral bone. . 129

10.4 3D mapping of mean absolute segmentation errors for femoral bone and cartilage. 130

10.5 3D mapping of maximal absolute segmentation errors for femoral bone and cartilage. 131

10.6 3D mapping of mean absolute segmentation errors for tibial bone. 132

10.7 3D mapping of maximal absolute segmentation errors for tibial bone. . . . 133

10.8 3D mapping of mean absolute segmentation errors for tibial bone and cartilage. 134

10.9 3D mapping of maximal absolute segmentation errors for tibial bone and cartilage. 135

10.10 Tissue contrast at the anterior femoral bone shaft. 137

10.11 Tissue contrast at the anteromedial border of the trochlea. 138

10.12 Tissue contrast at the tuberositas tibiae. 138

11.1 Methods 2 and 3: simulation of surgical navigation. 145

11.2 Methods 4 and 5: based on best-fit sphere of the femoral head 145

11.3 Comparison of measurement results in box-and-whisker diagram, split up for CAM and NO CAM cases. 146

11.4 Meniscal modeling 149

11.5 Meniscal geometry prediction and ground truth 150

11.6 Meniscal geometry prediction average error map 151

12.1 Methodology flowchart 157

12.2 Mean SMC and control group measurements. 161

12.3 Color map of the mean femoral shape distances between the SMC and control group. 164

12.4 Color map of the mean tibial shape distances between the SMC and control group. 165

12.5 Color map of the differences between a right SMC femur and the mean control femur 166

12.6 Color map of the differences between a right SMC tibia and the mean control tibia 166

13.1 Medial meniscus size distributions for donors and acceptors. 177

13.2 Lateral meniscus size distributions for donors and acceptors. 179

13.3 Estimated meniscus area donor versus acceptor. 181

13.4 Time to match in function of acceptor meniscus length. 181

14.1 CONSORT flow diagram for subject enrolment procedure. 190

14.2 3D visualisation of analysed bone region. 191

14.3 Schematic overview of the statistical shape analysis workflow. 192

14.4 KOOS subscores (mean ± standard deviation) for the MPMS group and R group 196

14.5 Shape variation in distal femur: comparison between R and MPMS group 197

14.6 Shape variation in proximal tibia: comparison between R and MPMS group 198

14.7 Shape variation in the tibiofemoral joint: comparison between R and MPMS group 199

14.8 Receiver operating characteristic curves for three distinct classifiers (distal femur, proximal tibia and tibiofemoral joint) 201

15.1 Schematic overview of the machine learning pipeline. 216

15.2 Learning curves for the logistic regression classifier. 219

15.3 ROC curves for the logistic regression classifier on the four non-overlapping test folds. 220

15.4 Shapley values for the 4-fold trained logistic regression models. 221

15.5 Principal component 0 of knee shape variation. 222

15.6 Principal component 2 of knee shape variation. 222

15.7 Principal component 3 of knee shape variation. 223

A.1 Bland-Altman plots for the automated versus manual measurements. . . . IX

List of Tables

7.1 Landmark definitions 77

7.2 Measurement definitions 79

7.3 Intraclass correlation coefficients for reliability analysis of the measurements. 84

7.4 Descriptive statistics for intra-observer, inter-observer and inter-method (manual versus automated) landmark position differences. 86

7.5 Descriptive statistics for measurement differences intra-observer, inter-observer, and intermethod (manual versus automated). 88

7.6 Intraclass correlation coefficients (manual versus automated) for reliability analysis of the measurements. 88

10.1 Conventional segmentation metric results from this experiment: mean +/- standard deviation (SD). 127

10.2 Conventional segmentation metrics for image segmentation. 135

12.1 Summary of the morphometric measurements for the femur and the tibia 162

13.1 Overall Differences Between Donor and Patient Populations. 177

13.2 Medial Meniscus Comparison of Meniscus Size, Anthropometric Factors, and Anatomic Side. 178

13.3 Lateral Meniscus Comparison of Meniscus Size, Anthropometric Factors, and Anatomic Side. 180

13.4 Effect of Sex and Anatomical Site on the Average Time to Match a Donor Graft. 182

14.1 Descriptive statistics for both R and MPMS group: patient sex, age, weight, height, BMI, cartilage status. 195

14.2 Correlation matrix of demographic, clinical, PROM- and SSM-derived variables. 200

15.1 Hyperparameter search space 215

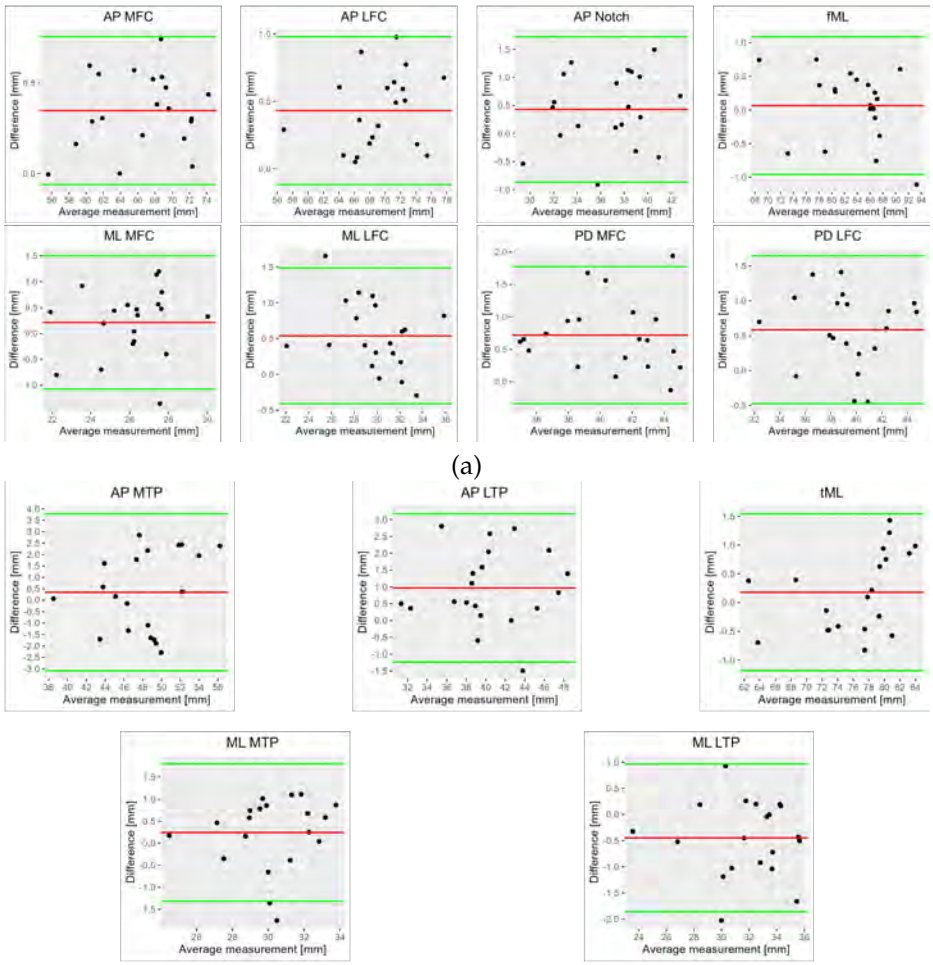
15.2 Descriptive statistics for clinical subpopulation variables. 217

15.3 Predictive model results following hyperparameter optimization (validation dataset, inner cross-validation loop). 218

15.4 Final predictive model results: generalizability to unseen data 218

Appendix A

Appendix



(a)

(b)

Figure A.1: Bland-Altman plots for the automated versus manual measurements (see Section 9.3.2 Automated Measurement Validation) of the **(a)** distal femur and **(b)** proximal tibia.

Appendix B

Curriculum vitae

Appendix C

List of Publications

C.1 Articles

- Grammens, J., Van Haver, A., Danckaers, F. et al. Small medial femoral condyle morphotype is associated with medial compartment degeneration and distinct morphological characteristics: a comparative pilot study Knee Surg Sports Traumatol Arthrosc. 2021 Jun;29(6):1777-1789.
- Van Oevelen, A., Duquesne, K., Peiffer, M., Grammens, J., Burssens, A., Chevalier, A., Steenackers, G., Victor, J., Audenaert, E. Personalized Statistical Modeling of Soft Tissue Structures in the Knee. Front Bioeng Biotechnol. 2023 Mar 8;11:1055860.
- Tabbaa, S.M., Pace, J.L, Frank, R.M., Grammens, J., Verdonk, P. An analysis of factors that influence meniscal allograft size discrepancies between patient and donor populations. Arthrosc Sports Med Rehabil. 2023 Apr 21;5(3):e569-e576.
- Van Fraeyenhove B, Verhaegen JCF, Grammens J, Mestach G, Audenaert E, Van Haver A, Verdonk P. The quest for optimal femoral torsion angle measurements: a comparative advanced 3D study defining the femoral neck axis. J Exp Orthop. 2023 Dec 18;10(1):141.
- Grammens, J., Van Haver, A., Lumban-Gaol, I., Danckaers, F., Verdonk, P., Sijbers J. Automated Landmark Annotation for Morphometric Analysis of Distal Femur and Proximal Tibia. Journal of Imaging. 2024; 10(4):90.
- Grammens, J., Van Haver, A., Danckaers, F., Vuylsteke, K., Sijbers, J., Mahluf, L. et al. (2024) Three-dimensional bone morphology is a risk factor for medial postmeniscectomy syndrome: a retrospective cohort study. Journal of Experimental Orthopaedics, 11, e12090. https://doi.org/10.1002/jeo2.12090
- Barbier, J., Van Haver, A., Grammens, J., Verborgt, O. Glenoid inclination and version measurement techniques used for 3D preoperative planning in total shoulder arthroplasty. (journal paper, submitted at Journal of Anatomy)
- Grammens, J., Danckaers, F., Van Haver, A., Verdonk, P., Sijbers, J. Shaping predictive models for arthroscopic partial medial meniscectomy outcome: a machine learning approach using clinical and bony morphological data from the

MEFISTO Project. (journal paper, submitted to Computer Methods and Programs in Biomedicine)

- Grammens, J., Danckaers, F., Van Haver, A., Verdonk, P., Sijbers, J. From Dice-scores to anatomical relevance: a comparative study of four state-of-the-art neural networks for automated knee MRI segmentation. (journal paper, submitted to NEJM AI)

C.2 Conferences

- Grammens, J., Peeters, W., Van Haver, A., Verdonk, P. Morphometric characteristics of the trochleodysplastic knee: a landmark-based 3D analysis (conference paper, podium presentation at EORS 2019, Maastricht, the Netherlands)
- Grammens, J., Van Haver A., Verdonk, P., Audenaert, E. Shape modelling for routine clinical practice in hip, knee and ankle pathology. (symposium proposal accepted for EORS 2021, Rome, Italy)
- Grammens, J., Tabbaa, S., Pace, J., Frank, R., Vanlommel, J., Van Haver, A., Verdonk, P. An analysis of medial meniscal allograft size discrepancies between patient and donor populations. (poster exhibition at ICRS 2022, 12-15 April 2022, Berlin, Germany).
- Grammens, J., Van Haver, A., Danckaers, F., Sijbers, J., Verdonk, P. Morphology of the distal femur and outcome of arthroscopic partial meniscectomy: Is there a link? (conference paper, podium presentation at ICRS 2022, 12-15 April 2022, Berlin, Germany).
- Grammens, J., Danckaers, F., Van Haver, A., Sijbers, J., Verdonk, P. Can we predict the future? – Morphological profiling. (invited speaker at morning workshop Meniscus Discovery Project, ICRS 2022, 12-15 April 2022, Berlin, Germany).
- Grammens, J., Tabbaa, S., Van Haver, A., Pace, J., Frank, R., Verdonk, P. An analysis of medial meniscal allograft size discrepancies between patient and donor populations (conference paper, poster exhibition at ESSKA 2022, 27-29 April 2022, Paris, France).
- Grammens, J., Van Haver, A., Danckaers, F. Sijbers, J., Verdonk, P. Morphology of the distal femur and outcome of arthroscopic partial meniscectomy: Is there a link? (conference paper, podium presentation at ESSKA 2022, 27-29 April 2022, Paris, France).
- Grammens, J., Danckaers, F., Van Haver, A., Sijbers, J., Verdonk, P. Do we know the future? (invited speaker at partner symposium ICRS during ESSKA 2022, 27-29 April 2022, Paris, France)
- Grammens, J., Danckaers, F., Van Haver, A., Sijbers, J., Verdonk P. Bone morphology and outcome in medial meniscectomy. (invited speaker at symposium "Bone Morphology in Knee Surgery" during 23rd EFORT Congress, 22-24 June 2022, Lisbon, Portugal).

- Grammens, J., Van Haver, A., Danckaers, F., Sijbers, J., Verdonk, P. Morphology of the distal femur and outcome of arthroscopic partial meniscectomy: Is there a link? (conference paper, pre-recorded electronic poster at The Meniscus Conference Luxembourg 7-9 July 2022).
- Grammens, J., Van Haver, A., Danckaers, F., Linder-Ganz, E., Angele, P., Kon, E., Sijbers, J., Verdonk, P. Can identification of bone morphology help us predict the future of our knees? A statistical shape modelling approach. (invited speaker at The Meniscus Conference Luxembourg 7-9 July 2022).
- Grammens, J., Van Haver, A., Danckaers, F., Sijbers, J., Verdonk P. Distal femur morphology and outcome of arthroscopic partial medial meniscectomy: development of a predictive algorithm. (Invited speaker at Orthopaedica Belgica, 27-28 April 2023, Brussels, Belgium)
- Grammens J., Pereira L.F., Danckaers F., Vanlommel J., Van Haver A., Verdonk P., et al. Accuracy evaluation of automatic segmentation in biomedical images: 3D spatial mapping of segmentation errors. Orthop Procs. 2024 Jan 2;106-B(SUPP_1):47-47. https://doi.org/10.1302/1358-992X.2024.1.047 (invited speaker at "Shape modelling for orthopaedic clinical research" symposium during EORS 2023, 27-29 September 2023, Porto, Portugal)
- Grammens J., Danckaers F., Van Haver A., Sijbers, J., Verdonk P., At the intersection of Sports, Prevention and Revalidation: Advanced 3D analysis to optimize success rates in knee surgery (invited speaker at "The Summit" by Flanders Technology & Innovation, 20-21 March 2024, Ghent, Belgium)
- Grammens, J., Van Haver, A., Danckaers, F., Sijbers, J., Verdonk P. Shaping the future: a predictive model for arthroscopic partial medial meniscectomy (invited speaker at Combined Belgian & Dutch Knee Society Meeting, 27 September 2024, Breda, the Netherlands)