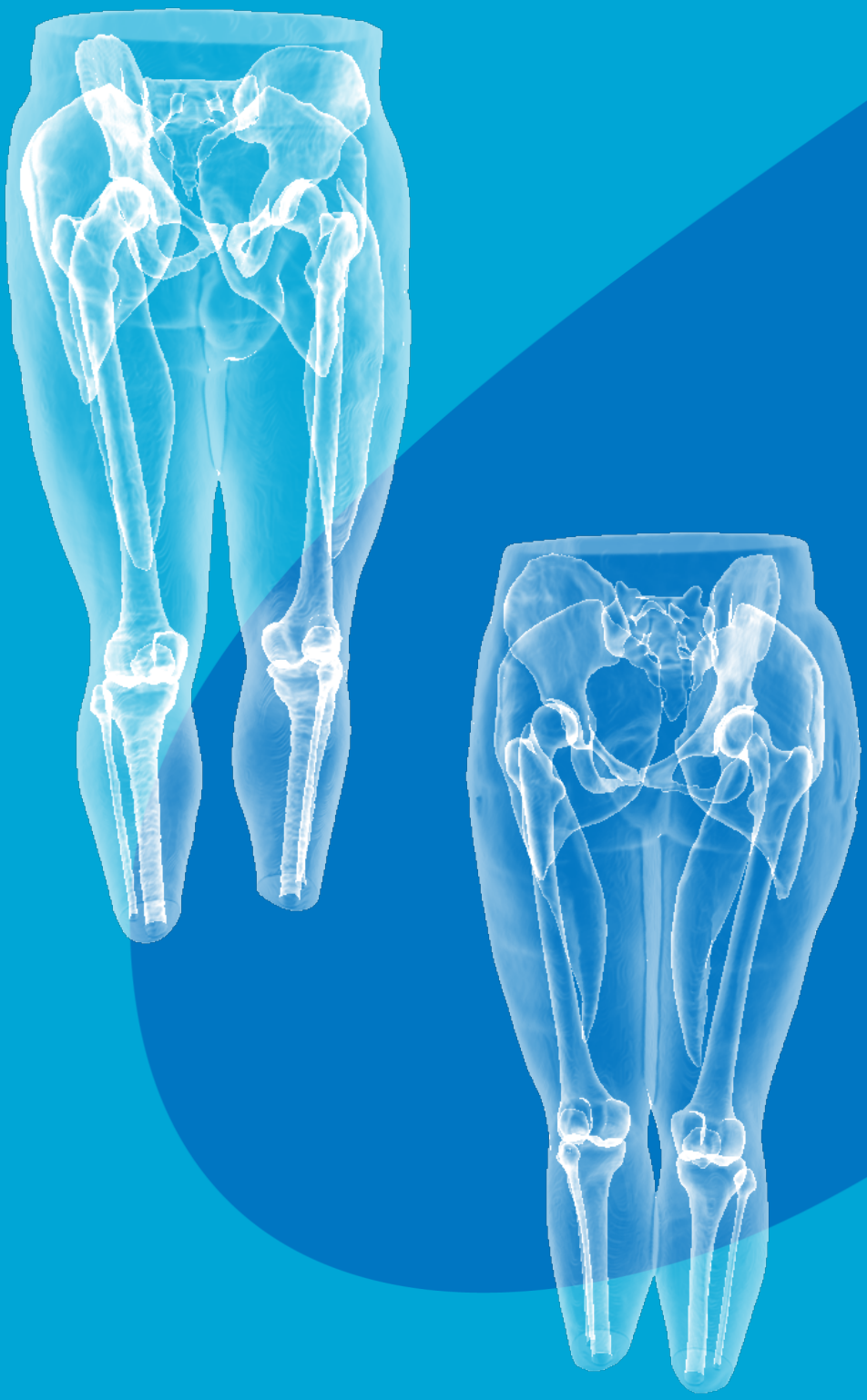
Development of an Automatic MRI-based Segmentation Model to analyze Sex-Related Differences in the Lower-extremity Musculoskeletal System

W.F. van de Meerakker

Master Thesis | MSc Biomedical Engineering | March 2025



TU Delft

Development of an Automatic MRI-based Segmentation Model to analyze Sex-related Differences in the Lower Extremity Musculoskeletal System

by

W.F. van de Meerakker

in partial fulfilment of the requirements to obtain the degree of

Master of Science
in Biomedical Engineering

at the **Delft University of Technology**,
to be defended publicly on Tuesday March 11, 2025 at 10:45 AM.

Student number:	4541332	
Project duration:	June, 2024 – March, 2025	
Thesis committee:	Ir. J. Cueto Fernandez,	TU Delft, daily supervisor
	Dr. ir. E. van der Kurk,	TU Delft, supervisor (chair)
	Prof. Dr. H.E.J. Veeger,	TU Delft, committeemember
	Dr. J. Hirvasniemi,	TU Delft & Erasmus MC, committeemember

TU Delft

Preface Handing in this thesis feels surreal, as it also marks the end of my time as a student. My journey—from starting in Industrial Design to deep-learning coding for this thesis—has made me realize how much I have grown. The past years have been challenging, with personal struggles along the way, but I am truly grateful for everything I have learned. Choosing Biomedical Engineering has fueled my passion for using technology to improve lives while deepening my fascination with human anatomy.

This achievement would not have been possible without the support of many. First, I sincerely thank my supervisors, Eline and Judith, for their guidance, encouragement, and inspiring discussions on reducing sex bias in biomechanical research. Their unwavering support, both academically and personally, has helped me push beyond my limits.

To my parents, thank you for your endless love and encouragement, reminding me to pursue what I truly enjoy rather than rushing to finish. To my friends, who have seen me at my highs and lows—your support has meant the world to me. A special thanks to Bernice, my roommate, whose insights helped me tackle challenges, and who always lent a listening ear. Lastly, to Céline, my fellow student and now a dear friend—our conversations have been inspiring, funny, and heartwarming, and I will truly miss them on a daily basis.

W.F. van de Meerakker
Delft, February 2025

***Abstract*—Despite the clear influence of sex on biomechanics, research continues to exhibit a sex bias, positioning male anatomy as the default. Consequently, current open-source musculoskeletal (MSK) models are primarily based on male bone geometry, mixed-sex musculotendon parameters and deviated from female muscle mass distribution. This raises concerns about the validity of MSK model predictions for females. While sexual dimorphism in the femur and pelvis is well-documented, its role in explaining sex-related differences in proportional muscle volumes remains unclear. This study aimed to investigate whether sex-specific variations in bone geometry can predict differences in the proportional volumes of the Gluteus Maximus (GMAX) and Rectus Femoris (RFEM), two muscles with known sex-based distinctions.**

To address this question, an automatic MRI-based segmentation model was developed using an nnU-Net deep-learning network trained on MRI scans of 16 healthy young adults (9 females, 7 males). Bone geometrical measurements were extracted using the STAPLE toolbox, and muscle volumes were normalized using lean volume to derive proportional volumes (%GMAX and %RFEM). Finally, linear stepwise regression was performed to assess whether bone metrics could predict muscle volumes.

The developed segmentation model demonstrated superior accuracy, outperforming existing methods with Dice Similarity Coefficients (DSC) of 0.926 for bones and 0.954 for muscles. Results revealed significant sexual dimorphism in bone geometry, with males exhibiting larger femoral offsets and knee widths, while females displayed greater posterior pelvic width and pelvic depth. %RFEM volume was significantly higher in males (p = 0.01), whereas %GMAX showed no significant sex-related differences. Regression analysis identified femoral offset and femur length as partial predictors of muscle volume proportions for %RFEM (R^2 = 0.478) and pelvis-femur length for %GMAX (R^2 = 0.151), but the low R^2 values indicate limited predictive power, particularly for %GMAX. These findings highlight potential errors in the calculation of proportional muscle volumes using lean volume, as well as limitations in the method used to measure bone metrics, suggesting that %RFEM and %GMAX cannot be reliably estimated using the included bone metrics.

These findings suggest that sexual dimorphism in the femur may partially explain sex-related differences in %RFEM, while its influence on %GMAX remains inconclusive. Further research is needed to fully understand the biomechanical implications of these differences. Integrating additional factors, such as muscle insertion points, actual muscle volume proportions, and a broader range of muscles to explore muscle leverage patterns, will help us understand the relationship between sexual dimorphism and muscle volume distribution better. As a result, this may improve the accuracy of sex-specific biomechanical simulations and deepen our understanding of sex-related differences in biomechanics.

I. Introduction

Until the 1990s, medical and scientific research was conducted almost exclusively on male subjects, as women were considered as "smaller men", apart from their reproductive system. However, studies in recent decades have shown significant sex-based differences in physiology and anatomy [1]. These findings have led to critical insights, such as the reduced efficacy or even harm of treatments tested solely on males when applied to women [2]. Moreover, the most common musculoskeletal diseases are more prevalent in females, including osteoporosis and osteoarthritis [3]. Sex-related differences also influence injury risk, with women being four to six times more likely to suffer anterior cruciate ligament (ACL) injuries in elite sports compared to men [4]. This discrepancy is linked to skeletal and musculotendon differences, which affects biomechanics among sexes [5].

Despite the evident role of sex in biomechanics, research continues to exhibit a sex bias, positioning male anatomy as the golden standard, a phenomenon referred to as bias-mechanics [6]. One of the primary tools for studying human movement is musculoskeletal (MSK) modeling, which embodies mathematical representations of the muscular and skeletal system [7]. While MSK models have the potential to improve our understanding of sex-related differences in biomechanics, they must first accurately reflect anatomical reality [8]–[10]. However, current MSK models remain male-based or rely on mixed-sex musculotendon parameters [11], [12]. A recent systematic review further revealed that female muscle mass distribution is not adequately represented in these models [13]. This questions the validity of MSK model predictions for females and underscores the need for a deeper understanding of sex differences in the musculoskeletal system.

Developing a scaling method that accounts for sex-related differences in muscle volume distribution would be a key step toward creating more accurate MSK models that reflect sex-specific variations. Notably, the Gluteus Maximus (GMAX) and Rectus Femoris (RFEM) show the most pronounced differences between both sexes, with RFEM proportional volumes being 12% higher in males and GMAX 17% higher in females [13]. These deviations are particularly relevant as proportional muscle volumes influence muscle leverage, significantly impacting estimated joint moments and muscle forces in static optimization during MSK modeling [14].

Research indicates a strong association between bone geometry and muscle volume [15], [16]. For example, femoral offset and the inter-femoral head distance are significant predictors of Gluteus Medius muscle volume across a diverse population with balanced sex distribution [16]. However, to our knowledge, no studies have investigated whether lower-extremity bone geometry can explain sex-related differences in proportional muscle volumes. Given that bones such as the pelvis have often been demonstrated to exhibit sex-based morphological differences [17], it is crucial to examine whether sexual dimorphism in the pelvis and femur contributes to variations in GMAX and RFEM proportional volumes. Understanding these relationships could provide valuable insights into sex-specific musculoskeletal adaptations and enhance the accuracy of MSK models.

The aim of this study is to investigate the relationship between sexual dimorphism in the pelvis and femur and the proportional volumes of the Gluteus Maximus (GMAX) and Rectus Femoris (RFEM). To achieve this, an automatic segmentation model was trained to extract the lower-extremity skeleton and relevant muscles from MRI scans from a cohort of 16 healthy young adults. Bone geometrical measurements were then derived from segmented bones using STAPLE [18] and analyzed for sexual dimorphism. Similarly, muscle volumes of GMAX and RFEM were examined to see if they

exhibited sex-related differences and compared to existing literature. Finally, relationships between bone geometry and muscle volume were explored to determine whether sexual dimorphism in the pelvis and femur could explain differences in proportional RFEM and GMAX volumes, potentially serving as a predictive tool for muscle volume distribution.

II. Methods

A. Participants

This thesis analyzed data from nine young adult female subjects (age: 28 $\pm$ 3 years, height: 172.4 $\pm$ 4.3 cm, weight: 63.8 $\pm$ 5.7 kg) and seven male subjects (age: 30 $\pm$ 4 years, height: 180.8 $\pm$ 10.6 cm, weight: 74.0 $\pm$ 10.1 kg). The MR images were obtained as part of a data collection effort for a PhD project within the BODIESlab at the BioMechanical Engineering department of TU Delft. Body weight and body fat percentage were measured using bioelectrical impedance analysis (BIA). This study exclusively included healthy individuals. For detailed anthropometric data for each subject, see the table in Appendix B.

B. MRI acquisition

All MRI scans were conducted by professionals at HollandPTC in Delft using a Philips Ingenia 3 Tesla scanner retrieving T1-weighted Dixon in-phase and water MR images for all subjects. The full body was imaged using 9 or 10 sequences in the superior-inferior (S-I) direction, depending on the subjects height. This study considered only stacks that capture the lower extremities, including the pelvic region and both legs. The majority of the sequences was captured with dimensions of 960 x 400 x 960. Still, the last two sequences, located around the lower legs and feet exhibited dimensions of 1024 x 400 x 1024 to capture the feet completely.

C. Post-processing of MRI sequences

To establish complete lower-body MR images, the seven sequences capturing the pelvic area, legs and feet, were registered using the NumPy [19] and NiBabel [20] libraries in Python. To address variations in voxel spacing and dimensions in the scans, all sequences were resampled to match the voxel dimensions of a reference sequence, typically the first sequence located around the pelvic area. This sequence was chosen as the reference because most sequences exhibited comparable voxel spacing, precisely 0.5729 [mm] x 0.7500 [mm] x 0.5729 [mm]. Only the last two sequences, located around the lower legs and feet, exhibited different voxel spacings. Empty slices and slices containing overlapping data from subsequent sequences were identified and removed from the preceding sequence to ensure accuracy in the concatenation process. Afterward, the sequences were concatenated in the inferior-to-superior direction based on their respective z-offsets to create full-leg compressed NIfTi files. Eventually, to reduce the computation time for the automatic segmentation models and enhance their model training by doubling the sample size, the full-leg MR images were also split symmetrically along the x-axis in Python to create an image for each leg. If the automatic splitting process was inaccurate, e.g. if it included muscle or bone segments from the opposite leg or pelvic region, manual cropping was performed in 3D Slicer [21]. The middle of the sacrum and the pubic symphysis were anatomical reference points for these manual corrections.

D. Segmentation of MRI data

This study focuses on the analysis of the lower limbs. Therefore, only bones and muscles in this region of interest were considered for segmentation. For manual segmentation, visualization, and refining automatic segmentations of the MR images, the biomedical imaging software 3D Slicer [21] was used.

1) Bone segmentation: For bone segmentation, all bones in the pelvic area, legs, and feet were divided into 19 separate segments in total and 10 segments per leg, where the sacrum was split in half (see Figure 12 in Appendix C). Since foot joints will not be modeled, the cuboid, cuneiforms, metatarsals, and phalanges were grouped into a single segment: the foot bones. On the other hand, the ankle joint was used in the MSK model and therefore, the calcaneus, talus and navicular were segmented into separate segments. In appendix C, an overview of the bones segmented per subject were given. TotalSegmentator, a nnU-Net-based deep-learning network trained on 561 MR scans across different modalities, was utilized for 4 subjects [22]. Although segmentation for some bones, e.g. the femur and pelvis was accurate, manual refinement was performed in 3D Slicer to improve precision. Additionally, certain bones, including the patella and foot bones, were not segmented by TotalSegmentator, at the time, and were therefore segmented entirely manually.

2) Muscle Segmentation: For muscle segmentation, various automatic segmentation models were explored. Among these, another trained 3D nnU-Net network [23], was employed to predict muscle structures for the muscles of interest, namely GMAX and RFEM. This 3D nnU-Net network, also referred to as the Henson099 model, had been trained on an open-source database comprising 69 MRIs and segmentations of 34 muscles in the lower limb of postmenopausal women [24] [25]. However, due to differences in imaging modalities and demographic characteristics between the training dataset and the MRI data used in this study, the predictions were suboptimal. Consequently, manual adjustments were made to the muscle segmentations to ensure accuracy.

3) Training of nnU-Net: Currently, U-Net-based architectures, a type of convolutional neural network used for semantic segmentation, are considered state-of-the-art in automatic segmentation [26], [27]. Therefore, nnU-Net [28] was employed in this study to train two separate networks for MRI segmentation: BODIES_skeletonLL047 for bone segmentation and BODIES_muscles021 for muscle segmentation.

An iterative training approach, consisting of two phases, was applied to reduce the time required for manual segmentation. See Figure 1 for a flowchart illustrating the training process. For the first training phase, 4 subjects were segmented using the semi-automatic approach described in previous subsections

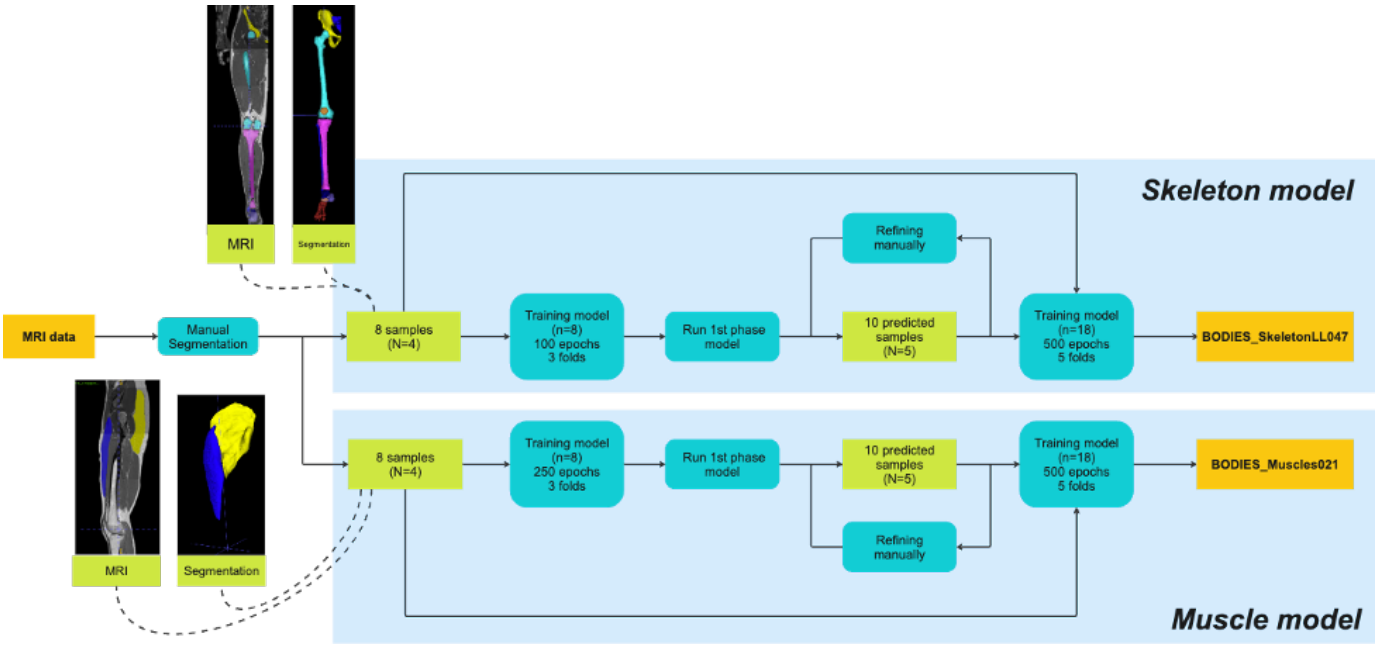


Fig. 1: Flowchart illustrating the training process for both nnUNet models, where *N* represents the number of subjects and *n* denotes the number of legs used as training sample.

II-D1 and II-D2. To increase the sample size for training the network and enhance computational efficiency, subject data was divided unilaterally, training each leg separately. This resulted in eight samples for the first training iteration (two per subject, one for each leg). For this training phase, networks for bone segmentation were trained on 3 folds using 100 epochs for 2 configurations, namely the 2D and 3D full resolution. After analyzing the results for bone segmentation, we choose to run the segmentation network for the muscle segmentation only for 3D full-resolution, as this delivered the best results. Furthermore, since the losses for the training and validation set were still high, it was decided to run it on 3 folds using 250 epochs. Subsequently, 5 additional subjects (again divided into 2 samples, capturing each leg) were run through the preliminary nnU-Net. Predictions were refined manually in 3D Slicer if needed. In the second and last training phase, the final nnU-Net for both models, namely the BODIES_skeletonLL047 and BODIES_muscles021, was trained on 18 samples. Each segmentation model's final nnU-Net was trained for the 3D full-resolution configuration across five folds, utilizing 500 epochs. Across all folds, different training/validation splits of 80/20 ratio were used, resulting in 14 samples for training and 4 for validation. Training was conducted in parallel on the DelftBlue supercomputer [29] using five NVIDIA A100 GPU nodes, each equipped with 80 GB of video RAM, and required around 37 hours to complete per fold. After training, cross-validation across the five folds was applied to remove bias introduced by the training set during the post-processing of predictions.

E. Data analysis

During data analysis, the final automatic segmentation models were validated against current automatic models from literature to ensure their accuracy. Afterwards, the relationship between bone geometrical measurements and the muscle volume ratios was analyzed.

1) Validation of automatic segmentation models: For analysis, one sample that was not in the training set were considered for validation. To analyze the robustness of the automatic segmentation models, the outcomes of these models were related to the ground truth, being the manual segmentation, using the DICE similarity coefficient (DSC) and the Hausdorff Distance (HD). The DSC is a common metric to evaluate segmentation performance by measuring the overlap between the ground truth and automatic segmentation, in the following way:

$$DSC = \frac{2 \cdot |A_{GT} \cap A_{Pre}|}{|A_{GT}| + |A_{Pre}|} \tag{1}$$

where $|A_{GT} \cap A_{Pre}|$ represents the number of overlapping voxels between the ground truth segmentation (A_{GT}) and the predicted segmentation (A_{Pre}). $|A_{GT}|$ is the total number of voxels in the ground truth. $|A_{Pre}|$ is the total number of voxels in the predicted segmentation. The DSC ranges from 0 to 1, where 1 indicates a perfect match between the predicted segmentation and the ground truth, while 0 indicates no overlap.

Since DSC is sometimes lacking in evaluating the true precision of the models, the Hausdorff distance was also analyzed. The HD measures the worst-case boundary error by computing the maximum distance between the ground

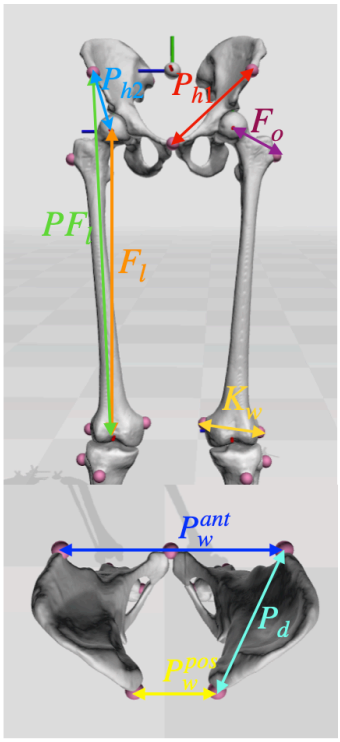


Fig. 2: Bone measurements illustrated that were obtained using STAPLE [18].

truth and the predicted segmentation surface, expressed in millimeters:

$$HD = \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(y,x)\right\} \quad (2)$$

where X represents the set of points on the surface of the ground truth. Y represents the set of points on the surface of the predicted segmentation. $d(x,y)$ is the Euclidean distance between points. The function $\inf$ finds the closest point distance, and $\sup$ ensures the maximum of all minimum distances.

Both metrics were calculated using the SciPy [30] and NumPy [19] libraries in Python. Both metrics were calculated for the separate segments as well as for the complete segmentation predicted by the models. A heatmap of the HD was created using MeshLab [31] to get more insight on location of these errors.

2) Analysis of bone geometry and muscle volumes: The analysis of bone geometry and muscle volumes was conducted on data collected unilateral, yielding a total of 32 samples, consisting of 18 samples from female subjects and 14 samples from male subjects.

a) Calculation of bone geometrical measurements: To get insights on the bone geometry, the partial bone segmentations were transformed into a skeletal model of the lower extremity using the STAPLE toolbox [18] in MATLAB R2024b [32]. The STAPLE toolbox makes use of algorithms

Variable	euclidian distance between
Femur length (F_l)	knee joint center and the hip joint center
Femoral offset (F_o)	hip joint center and the greater trochanter
Knee width ($Knee_w$)	Lateral epicondyle and medial epicondyle of the femur
Pelvis height (P_{h1})	pubic symphysis and ASIS
Pelvis height (P_{h2})	hip join center and ASIS
Pelvis-Femur length (PF_l)	ASIS and knee joint center
Pelvis depth (P_d)	ASIS and PSIS
Pelvis width (P_w^{ant})	left and right ASIS
Pelvis width (P_w^{pos})	left and right PSIS

TABLE I: Description of the bone geometrical measurements and their derivation based on particular anatomical landmarks.

to fit spheres and cylindrical structures on the bone segmentations, identifying joint centers and marker placements on bony landmarks. Additionally, the centers of the bone segment were explored. Euclidean distances between data points were computed to derive the bone geometrical measurements related to pelvic and femoral morphology, listed in Table I.

Figure 2 illustrates these metrics on the bilateral model created by STAPLE. Additional metrics were initially considered but excluded from the results due to measurement errors or the statistical criteria outlined in Section II-E2c. The specific reasons for exclusion, along with a complete list of these metrics, are provided in Appendix D. Furthermore, bone metrics were normalized by body height to enable a fair comparison when evaluating sexual dimorphism.

b) Calculation of Muscle volumes: Muscle volume for each leg was determined by the number of voxels per muscle segment and multiplying it by the voxel volume, derived from the voxel spacing in the segmentation. The muscle volume formula is:

$$V_{muscle} = N_{voxels} * V_{voxel} * 10^3$$

where V_{muscle} is the muscle volume in $[cm^3]$, N_{voxels} is the number of voxels and V_{voxel} is the volume of one voxel in $[mm^3]$.

Since the total muscle volume in the lower limbs was not segmented, muscle volumes were normalized by their partition in the lean volume of the lower limbs. Total volume of the lower limbs was derived by automatic segmentation of the body by TotalSegmentator. This volume was then cropped just above the pelvis and above the ankle joint, at the distal end of the tibia and fibula shafts before widening begins, since MRI data for the feet was missing for one subject. Skeletal structures within this cropped volume were subtracted, yielding the soft tissue volume:

$$V_{soft} = V_{body} - V_{skeleton}$$

where V_{soft} is the volume of all soft tissues in this cropped volume in $[cm^3]$, V_{body} is the total body volume $[cm^3]$ and $V_{skeleton}$ is the volume of bones within this cropped volume in $[cm^3]$.

To derive lean volume, fat volume had to be derived first. Since fat has a lower density than total of all soft tissues, the proportional fat volume within soft tissue was corrected using the density ratio between lean soft tissue and fat tissue:

$$\%V_{fat} = \frac{\rho_{soft}}{\rho_{fat}} * \%M_{fat}$$

$$V_{fat} = V_{soft} * \frac{\%V_{fat}}{10^2}$$

where ρ_{lean}, which is 1.10 $[g/cm^3]$ represents the mass density of lean soft tissue (excluding bone tissue) and ρ_{fat} is the mass density of fat tissue which is 0.9 $[g/cm^3]$ [33]. $M_{fat}\%$ refers to the percentage of fat in the body mass. V_{fat} is the fat volume in $[cm^3]$, whereas $V_{fat}\%$ represents fat as a percentage of soft tissue volume.

Afterwards the lean volume (V_{lean}) was calculated in $[cm^3]$ by substracting fat volume from the soft tissue volume:

$$V_{lean} = V_{soft} - V_{fat}$$

Using the lean volume, the proportional muscle volumes ($V_{muscle}\%$) were derived in the following way:

$$\%V_{muscle} = \frac{V_{muscle}}{V_{lean}} * 10^2$$

c) Statistical analysis: Data for bone geometry and muscle volume were first tested for normality using the Shapiro-Wilk Z-test and for homogeneity of variance using Levene's test. Paired t-tests were conducted to assess bilateral symmetry in bone geometry measurements and muscle volumes between legs. Students t-tests were used to analyze sex-related differences in (proportional) muscle volumes and bone metrics when Levene's test confirmed homogeneity of variance; otherwise, a Welchs t-test was performed.

To identify the most predictive bone geometrical parameters for muscle volume proportions, linear stepwise regression was employed. Significantly asymmetrical bone metrics between legs were excluded from regression to account for measurement error. In addition to bone metrics, anthropometric variables such as body height, body mass, and sex were included as independent predictors.

This method used bidirectional selection, combining forward selection and backward elimination. In forward selection, variables were added if they minimized the Akaike Information Criterion (AIC), which evaluates model quality by balancing goodness-of-fit and model complexity. Additionally, variables were only retained if they significantly improved model performance ($p < 0.05$). In backward elimination, variables with $p > 0.10$ were removed to prevent overfitting and improve model interpretability.

The final Ordinary Least Squares models were evaluated using R^2 and adjusted R^2, with only statistically significant predictors ($p < 0.05$) reported. The best model was selected based on AIC minimization, ensuring a balance between model fit and complexity. Additionally, R^2 and adjusted R^2 were used to assess explanatory power, while only predictors with $p < 0.05$ were retained to ensure statistical significance.

All statistical analyses were conducted in Python, utilizing the Seaborn [34] and SciPy [30] libraries.

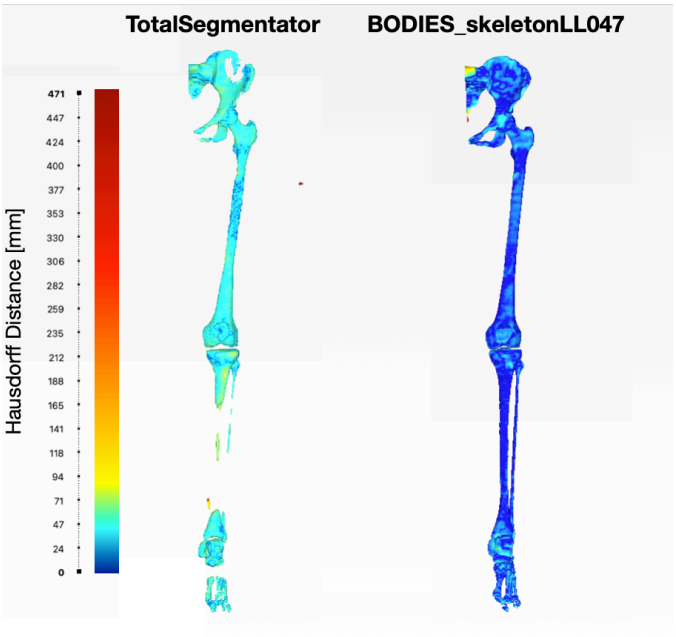


Fig. 3: Heatmaps illustrating location of the Hausdorff distances on the meshes of the prediction by both models for the lower extremity skeleton. Red areas indicate inaccurate predictions, whereas dark blue areas present better accuracy.

Bone	*BODIES_SkeletonLL047*		*TotalSegmentator*	
	DSC $[-]$	**HD** $[mm]$	**DSC** $[-]$	**HD** $[mm]$
Pelvis	0.900	15.78	0.770	38.86
Sacrum	0.851	55.64	0.831	20.00
Femur	0.960	11.47	0.908	12.25
Patella	0.848	8.77	0.881	8.36
Tibia	0.964	9.90	0.719	222.33
Fibula	0.884	15.39	0.570	210.83
Calcaneus	0.902	10.09	0.848*	16.40*
Talus	0.881	11.22	-	-
Navicular	0.871	8.77	-	-
Foot bones	0.801	16.31	0.387	1431.78
Overall	0.926	24.50	0.807	208.49

TABLE II: Results of the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) for the different automatic bone segmentation models on the sample used for validation. * is denoted as these values represent the DSC and HD for the Calcaneus, Talus and Navicular, since these were not separately segmented by TotalSegmentator.

III. Results

A. Validation of segmentation models

1) Bone segmentation models: Table II presents the Dice Similarity Coefficients (DSC) and Hausdorff Distances (HD) for individual segments and overall predictions by both models. At first glance, both models perform reasonably well, with DSC values ranging from 0.807 to 0.964. The BODIES_SkeletonLL047 model consistently outperformed TotalSegmentator, achieving higher DSC values overall. The HD results reveal distinct differences between the models. TotalSegmentator's overall HD remained exceptionally high

Muscle	*BODIES_Muscles021*		*Henson099*		*TotalSegmentator*	
	DSC $[-]$	**HD** $[mm]$	**DSC** $[-]$	**HD** $[mm]$	**DSC** $[-]$	**HD** $[mm]$
Gluteus Maximus	0.953	19.05	0.834	175.65	0.756	52.21
Rectus Femoris	0.960	17.11	0.636	25.73	-	-
Total	0.954	17.11	0.798	175.65	0.756	52.21

TABLE III: Results for the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) achieved by various automatic muscle segmentation models on the validation sample. Note: TotalSegmentator does not provide separate segmentation for the Rectus Femoris (RFEM).

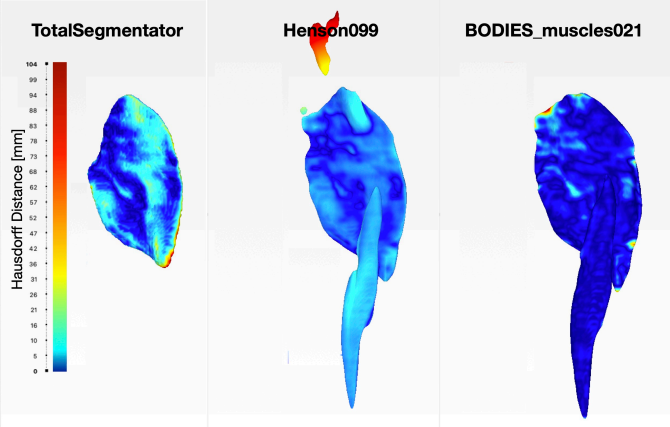


Fig. 4: Heatmaps illustrating location of the Hausdorff distances on the meshes of the prediction by all the evaluated models for the muscles. Note: TotalSegmentator did not segment RFEM.

(208.49 $[mm]$), whereas the BODIES_SkeletonLL047 model exhibited a significantly lower overall HD value of 24.50 $[mm]$.

Analysis of individual segments further highlights the superiority of the BODIES_SkeletonLL047 model. In particular, the Pelvis, Tibia, and Fibula were predicted with far greater accuracy compared to TotalSegmentator, as evidenced by Figure 15. The largest disparity was observed for the Fibula, where BODIES_SkeletonLL047 achieved a DSC of 0.884 and HD of 15.39 $[mm]$, compared to a DSC of 0.570 and HD of 210.83 $[mm]$ by TotalSegmentator. Similarly, the Pelvis and Tibia predictions by TotalSegmentator showed noticeable segmentation gaps (see Figure 15). Both models performed similarly for the Sacrum (DSC: 0.831 vs. 0.851) and Femur (DSC: 0.908 vs. 0.960), both achieving high performances for the Femur.

Notably, BODIES_SkeletonLL047 encountered modest challenges with smaller structures, such as the Patella and all bones in the feet, yet still outperformed TotalSegmentator in these regions. The Patella was predicted with a DSC of 0.848 and an HD of 8.77 $[mm]$ by BODIES_SkeletonLL047, compared to 0.881 DSC and 8.26 $[mm]$ HD by TotalSegmentator, demonstrating comparable performance for this bone. However, TotalSegmentator's results for the Foot Bones were substantially less accurate (DSC: 0.387, HD: 1431.78 $[mm]$), compared to BODIES_SkeletonLL047 (DSC: 0.801, HD: 16.31 $[mm]$), highlighting significant segmentation errors in these smaller structures. Additionally, as TotalSegmentator did not segment the Calcaneus, Talus, or Navicular separately, a combined DSC of 0.848 and HD of 16.40 $[mm]$ was reported for these bones, whereas BODIES_SkeletonLL047 demonstrated stronger performance with individual DSC values of 0.902, 0.881, and 0.871 for the Calcaneus, Talus, and Navicular, respectively.

In general, the BODIES_SkeletonLL047 model demonstrated better predictions than TotalSegmentator, particularly for the Tibia, Fibula, and smaller bones, while maintaining robust accuracy across most bone structures.

2) Muscle segmentation models: Two other models, Henson099 and again TotalSegmentator, were compared and analyzed against the trained model in this study, namely BODIES_Muscles021. Similar to the evaluation of bone segmentation models, muscle segmentation performance was assessed using the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) metrics, with results summarized in Table III.

BODIES_Muscles021 demonstrated the best overall performance, achieving the highest DSC values (0.953 to 0.960) and the lowest HD values (17.11 $[mm]$ to 19.05 $[mm]$) among the three models. Although Henson099 outperforms TotalSegmentator with higher DSC (0.834 vs. 0.756), its HD was significantly higher for Henson099 (175.65 $[mm]$ vs. 52.21 $[mm]$). Both Henson099 and TotalSegmentator showed volume shrinkage compared to ground truth, as evident in cross-sectional comparisons (see Appendix H).

A closer examination of muscle-specific performance reveals that Henson099 achieved notably higher DSC values for the GMAX (DSC: 0.834) than for the RFEM (DSC: 0.636), whereas HD is much higher for GMAX than RFEM, indicating suggested noise in specific areas (see Figure FIXME). In contrast, BODIES_Muscles021 achieved consistent performance across both muscles, with balanced DSC (GMAX: 0.953, RFEM: 0.960) as well as HD values (GMAX: 19.05 $[mm]$, RFEM: 17.11 $[mm]$).

Overall, BODIES_Muscles021 exhibited the most robust and reliable segmentation, consistently achieving high accuracy and minimal errors across both muscles.

B. Analysis of bone geometry and muscle volumes

1) Bone geometry measurements: Figure 5 presents violin plots of the analyzed bone metrics in this study, while Figure 7 illustrates the same values normalized by body height. Most femoral measurements were consistently larger in males, whereas pelvic measurements exhibited greater variability

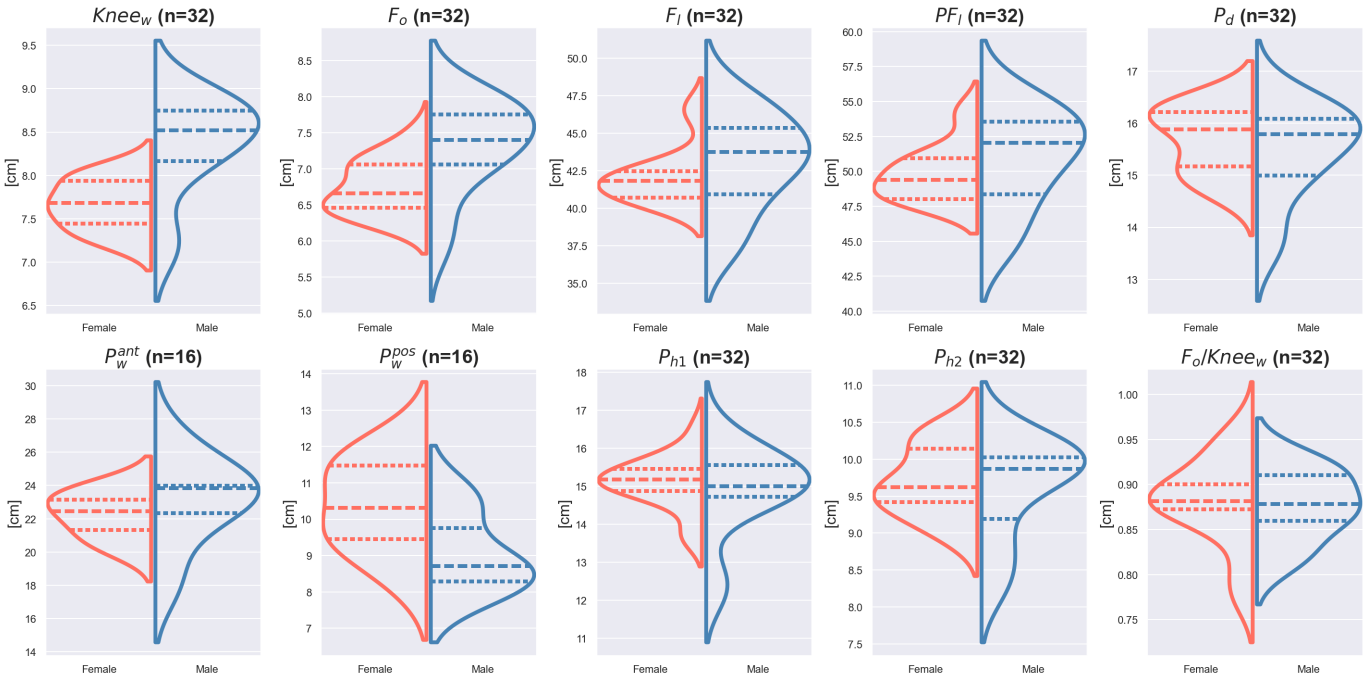


Fig. 5: Violin plots comparing bone metrics between female and male subjects. The broken dotted lines inside the plots represent the 1st and 3rd quartiles, while the middle dashed line indicates the median. *n* represents the number of available data points for each metric, which varies for some measurements due to bilateral assessment.

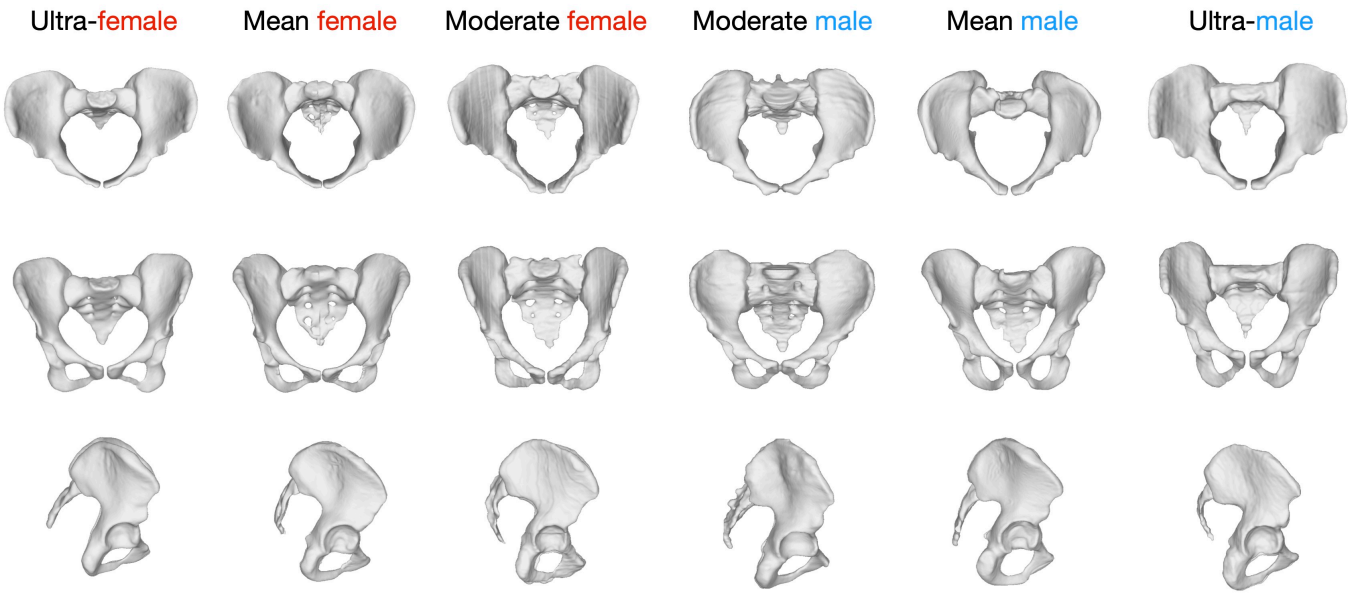


Fig. 6: Pelvic morphologies from our cohort ordered on their sexual dimorphism, based on the analysis by Fischer et al. (2017) [17]. From left to right, pelves transitioning from an 'ultra-female' to an 'ultra-male' morphology. The 'ultra-female' pelvis showed a more anterior-tilted orientation with a higher pelvic inlet and straight posteriorly pointing sacrum. In contrast, the 'ultra-male' pelvis features a more vertically aligned ilium, a narrower inlet, and an inwardly curved sacrum.

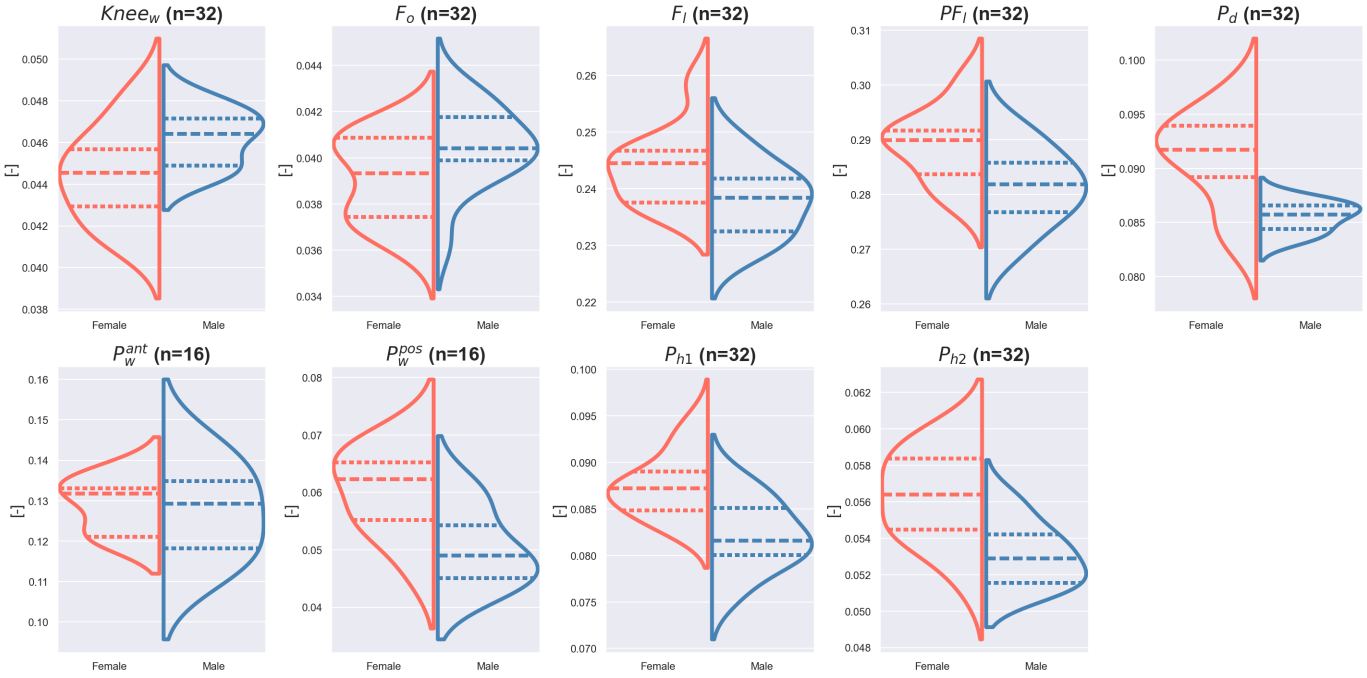


Fig. 7: Violin plots comparing bone metrics normalized by body height between female and male subjects. *n* represents the number of available data points for each metric, which varies for some measurements due to bilateral assessment. The ratio $F_o/Knee_w$ was excluded as it is inherently independent of body height.

between sexes. When normalized by body height, pelvic measurements were larger in females.

Posterior pelvic width was significantly greater in females ($p = 0.048$), whereas anterior pelvic width showed no significant difference between sexes ($p > 0.05$). These differences persisted even after normalization by body height. Although other pelvic dimensions did not exhibit significant sex differences, males displayed greater variability in absolute pelvic metrics, except for posterior pelvic width, where females showed higher variability. However, after normalization, nearly all pelvic measurements (except anterior pelvic width) were significantly larger in females ($p < 0.01$), with pelvic depth showing the most pronounced variability in females. In contrast, anterior pelvic width exhibited greater variability in males.

Among femoral measurements, knee width and femoral offset were significantly larger in males ($p < 0.001$), a difference that persisted after normalization ($p < 0.05$). When femoral offset was normalized by knee width, the ratio remained similar across sexes ($p > 0.05$), indicating proportionality between these two metrics. Additionally, femur length and pelvis-to-femur length did not significantly differ between sexes in absolute terms ($p > 0.05$), but after normalization, these values were larger in females ($p < 0.01$). Although absolute femoral measurements exhibited greater variability in males, after normalization, variability was comparable between sexes, except for knee width, where females had a slightly wider range.

Overall, femoral offset and knee width were consistently larger in males, whereas most normalized pelvic metrics, such as pelvic height, pelvic depth, and posterior pelvic width, were larger in females. Variability was similar for most metrics, though pelvic depth showed greater variation in females, while anterior pelvic width was more variable in males.

2) Muscle Volume: Table IV presents a comparison of muscle volumes measured in this study with values reported in the literature. To ensure a fair comparison, the referenced studies also investigated a young adult population, including female and male subjects, with muscle volumes derived from MRI. The study by Handsfield et al. (2014) [35] only published values for the total population without distinguishing between sexes.

Interestingly, values for both muscles fall within the range of values reported in previous literature. Whereas our cohort exhibited consistently higher muscle volumes than those found by Lube et al. (2015) [36], Handsfield et al. (2014) [35] reported overall higher muscle volumes. Similarly to Lube et al. (2015), muscle volumes for the GMAX and RF were higher in male subjects than in female subjects.

The violin plots in Figure 8 further illustrate these sex-based differences. The distribution of absolute muscle volumes in female subjects appears more uniform, while male subjects display a wider range, including some instances of higher muscle volumes. Notably, proportional muscle volume for %RFEM is significantly higher in males ($p = 0.01$), whereas for %GMAX, no significant difference was found ($p > 0.05$) between sexes based on the t-test.

		Female Average ± STD [cm^3]	Male Average ± STD [cm^3]	Overall Average ± STD [cm^3]
This study	*GMAX*	756.7 ± 118.6	990.6 ± 90.3	859.0 ± 158.2
(N=16, F/M: 9/7)	*RFEM*	194.2 ± 29.7	286.0 ± 45.4	234.3 ± 59.1
Handsfield et al. [35]	*GMAX*	-	-	849.0 ± 194.7
(N=24, F/M: 8/16)	*RFEM*	-	-	269.0 ± 64.3
Lube et al. [36]	*GMAX*	657.2 ± 115.1	871.1 ± 34.4	764.1 ± 138.0
(N=6, F/M: 3/3)	*RFEM*	158.9 ± 23.5	265.7 ± 30.8	212.3 ± 61.6

TABLE IV: Comparison of muscle volumes from this study with MRI-derived values reported in the literature for young adult cohorts. N: number of participants, F/M: partition between sexes.

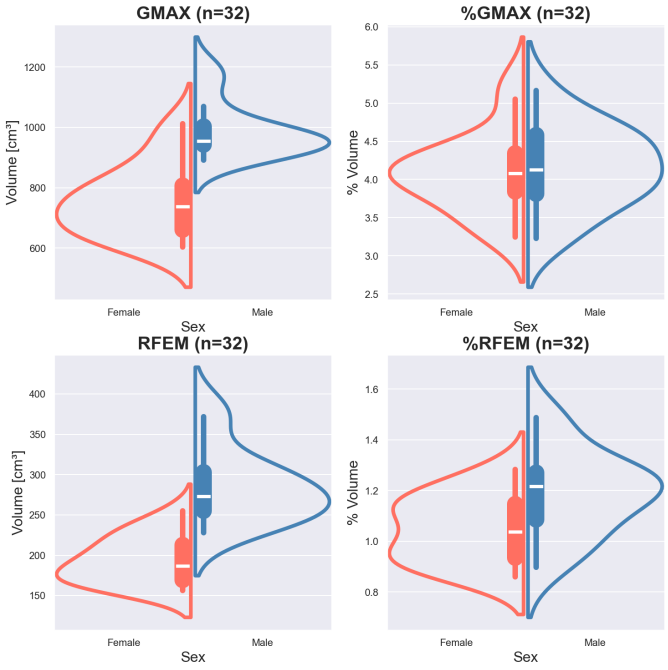


Fig. 8: Violin plots comparing muscle volumes [cm^3] and the proportions of muscle volume [%] within the lean volume between sexes. *n* represents the number of muscles, with one per leg.

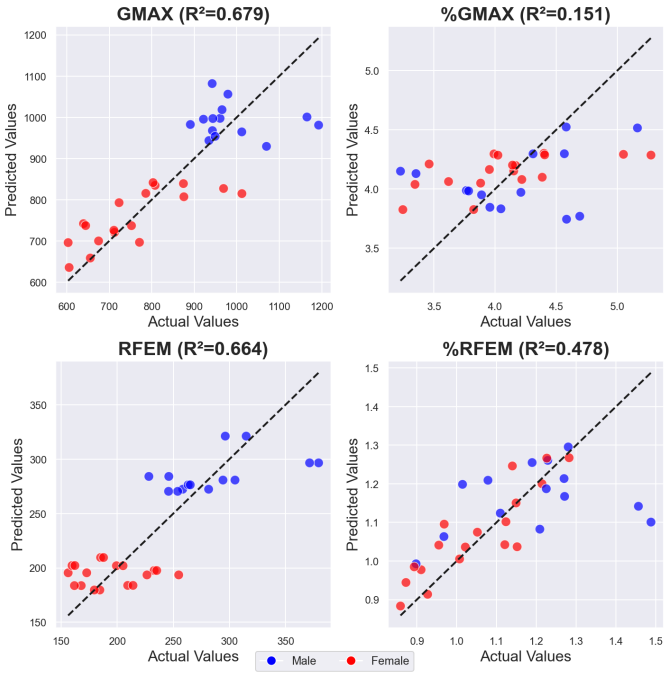


Fig. 9: Scatter plots showing predicted versus actual values for (proportional) muscle volumes. The dashed line indicates perfect prediction, where data points on the line correspond to accurate predictions. Red points represent female subjects, while blue points represent males.

3) Stepwise regression: Table V presents the regression equations for (proportional) muscle volumes, along with R^2, adjusted R^2 values, indicating the variance explained by the model, and the standard estimated error (SEE). Figure 9 visualizes the predicted vs. actual values, illustrating the accuracy of the regression equations.

Sex emerged as the strongest predictor ($p < 0.001$) for absolute muscle volumes (GMAX and RFEM) but was not a significant predictor for proportional volumes. Moreover, the models predicting absolute muscle volumes (GMAX: R^2 = 0.679, RFEM: R^2 = 0.664) demonstrated higher accuracy, explaining more variance compared to the proportional volume models (%GMAX: R^2 = 0.151, %RFEM: R^2 = 0.478). In particular, the low R^2 for %GMAX suggests poor predictability, which is further reflected in Figure 9; its scattered data points do not follow the regression trend. Notably, %GMAX had only one significant predictor, whereas the other models had at least two.

For GMAX and %GMAX, both pelvic and femoral metrics were predictive, indicating dependency on both bones. Interestingly, femur length and pelvis-to-femur length (capturing part of femur length) negatively correlated with GMAX and %GMAX, suggesting that a longer femur is associated with lower GMAX volume.

While RFEM had no bone metrics in its predictive equation, %RFEM was dependent on femoral metrics. Noteworthy, femur length also negatively influenced %RFEM, similar to its effect on GMAX and indirectly on %GMAX. Femoral offset was also predictive, positively affecting %RFEM.

IV. Discussion

In the first part of this thesis, an automatic segmentation method was developed for the segmentation of the lower extremity skeleton and two muscles (GMAX and RFEM) from T1-weighted MRI, such that bone and muscle data

Predicted muscle	Prediction equation	R^2	Adj. R^2	SEE
GMAX $[cm^3]$	$276.757 * Sex(F = 1, M = 2) + 98.817 * Pelvis_d - 22.131 * Femur_l - 137.596$	0.679	0.645	88.194 $[cm^3]$
%GMAX	$-0.074 * PF_l + 7.857$	0.151	0.122	0.475 [%]
RFEM $[cm^3]$	$73.898 * Sex(F = 1, M = 2) + 1.758 * BW + 8.193$	0.664	0.641	33.692 $[cm^3]$
%RFEM	$-0.075 * Femur_l + 0.280 * Femur_o + 2.332$	0.478	0.442	0.116 [%]

TABLE V: Regression equations for GMAX and RFEM (proportional) muscle volumes. All variables were statistically significant at p < 0.05. BW refers to body weight in $[kg]$, while all other variables, except Sex, represent bone metrics mentioned earlier.

could be derived. This was successfully achieved through training of nnU-Net, a deep-learning network, on 18 samples, training two separate networks: one for bone segmentation, i.e. the BODIES_skeletonLL047 model, and one for the muscles, known as the BODIES_muscles021 model. Both models demonstrated superior performances with a DSC of 0.926 for the BODIES_skeletonLL047 and a DSC of 0.954 for BODIES_muscles021, significantly outperforming existing segmentation models.

While the BODIES_skeletonLL047 model is no longer the first to segment all bones in the lower limbs from MRI, its ability to separately segment the Calcaneus and Talus makes it highly valuable for automatic pipelines in the generation of subject-specific MSK models of the lower extremity. MRI-based data collection may be preferred over CT-based methods as it avoids radiation exposure and, in the context of MSK modeling, allows for the simultaneous collection of both skeletal and muscle data in a single acquisition. While TotalSegmentator combines the Calcaneus, Talus, and Navicular into a single segment and exhibits low accuracy for foot bones, BODIES_skeletonLL047 provides more detailed segmentation of these structures. Additionally, it outperforms TotalSegmentator in key regions such as the Pelvis, Tibia, and Fibula, making it a robust choice for MSK model development.

However, performance varied across anatomical structures. Smaller bones, such as the sacrum, patella, and foot bones, exhibited lower DSC values compared to larger bones, indicating reduced segmentation accuracy. This discrepancy is expected, as DSC is a relative measure, and smaller structures inherently lead to greater proportional errors. Despite this, the HD values for the patella and foot bones were comparable to those of larger bones, suggesting that these structures were still segmented with high accuracy. The primary exception was the sacrum, which demonstrated the highest HD, indicating a notable segmentation error. A closer analysis of the segmentation heatmaps (Figure 3) revealed that errors primarily occurred at the base (sacral promontory) and tailbone (coccyx) of the sacrum. The tailbone segmentation error was likely due to inconsistencies in the ground truth rather than a model deficiency, as BODIES_skeletonLL047 appeared to capture the structure more accurately. In contrast, errors at the sacral base may have stemmed from training samples where parts of the first vertebral disc were mistakenly included, leading to over-segmentation in this case.

Likewise, BODIES_muscles021 has a very high performance for both muscles and outperforms the other models substantially. Although the model, for now, is only trained for two muscles, it demonstrates the power it has in segmenting muscles, even when accounting for variability in muscle morphology between subjectsparticularly for RFEM. Notably, some females exhibited a more curved and rounded RFEM, whereas in both males and some females, RFEM appeared straighter and flatter (see Appendix K).

Notably, the errors observed in the comparison models do not appear to stem from deficiencies in model architecture, as these models were also trained using nnU-Net. Instead, discrepancies likely arise from inconsistencies in segmentation quality, variations in MRI modalities, lower image resolution, and differences in the demographics of the training populations.

For instance, Henson099 was trained on a different MRI modality with lower image quality, and the segmentations used for training were less refined (see Appendix 4). These factors point out that muscle segmentation models are highly dependent on the imaging modality they are trained on. As a result, it cannot be guaranteed that BODIES_muscles021 will achieve the same high performance on MRIs acquired with different modalities or scanner settings (see Appendix I). Furthermore, Henson099 was trained on an elderly female population (mean age: 69 $\pm$ 7 years), which differs significantly from the demographic composition of this study's participant group. MRI scans from that dataset reveal increased subcutaneous fat and lower muscle volumes, which may influence segmentation performance when applied to younger or more physically active individuals. Conversely, it can not be confirmed whether the BODIES_muscles021 has promise in segmenting elderly muscles accurately.

Similarly, TotalSegmentator, while trained on a diverse range of MRI modalities, presents its own set of limitations. The broad applicability of this model is an advantage; however, segmentation accuracy was compromised to optimize computational efficiency. Additionally, although TotalSegmentator was trained on a large dataset (561 MRI scans), a significant portion of these scans focused on the abdominal region, rather than the lower extremities. Specifically, only 60 to 70 scans included the pelvic and hip region, and even fewer covered the legs, particularly the lower legs. Given this limited representation of lower-extremity structures, together with the reduced segmentation quality in the training set (see Appendix I), the model may not be as robust for detailed and accurate muscle and bone segmentation in these regions compared to our models.

Moreover, both Henson099 and TotalSegmentator were trained on lower-quality MRIs compared to the higher-resolution scans used in this study, further enhancing the precision of BODIES_muscles021 and BODIES_skeletonLL047 at a finer scale. This suggests that, despite TotalSegmentators adaptability across different MRI modalities, our models likely provide higher segmentation accuracy for the lower-extremity skeleton and the two muscles in high-quality T1-weighted MRI scans.

The second part of this study was to analyze sex-related differences in (proportional) muscle volumes and bone geometry. Consequently, if differences were found in the bone geometry, they could be related to, or even explain, the differences in the muscle volumes.

The analysis of bone geometrical measurements revealed that several femoral metrics were consistently larger in males. As expected, males demonstrated significantly greater femoral offset and knee width than females ($p < 0.003$). However, femur length did not differ significantly between sexes ($p > 0.05$), which contrasts with established literature associating longer femurs with greater body height, typically observed in males [35], [37]. Interestingly, when normalized by body height, femur length appeared proportionally greater in females ($p = 0.01$) (see Table XI in Appendix G). This suggests that females may have relatively longer femurs, though the measurement was taken between joint centers rather than along the femoral shaft. The result may be influenced by the femoral neck angle relative to the femoral shaft (FSA) in the frontal plane, indicating a higher FSA in females, which is supported by existing literature [38], [39]. Despite males having greater overall body height ($p = 0.003$), femoral offset and knee width remained significantly larger in males even after normalization ($p < 0.05$).

Notably, the ratio between femoral offset and knee width was nearly symmetrical between sexes (F: 0.881 vs. M: 0.876, $p > 0.05$), suggesting a proportional relationship between these two measures. This symmetry may reflect the role of femoral offset in determining knee-loading conditions and the adaptive response of bone in the knee joint to mechanical forces. Increased femoral offset has been shown to influence joint alignment and load distribution across the hip and knee, affecting muscle moment arms and joint reaction forces, as observed in total hip arthroplasty [40]. However, as previously mentioned, FSA was not accounted for in these measurements, limiting the interpretation of this relationship. Additionally, femoral anteversion, defined as the angle between the femoral neck and the knee joint center axis, reflects the degree of femoral torsion. This structural parameter plays a crucial role in joint mechanics and load distribution across the hip and knee. Given its influence on the position of the femoral neck relative to the knee, it may also be relevant for analyzing its influence regarding this ratio [41]. Despite the absence of these factors, the observed proportionality between femoral offset and knee width suggests that sex-related differences in knee-loading conditions may originate from sexual dimorphism in the femoral neck and head. If this ratio remains consistent across larger populations, including different age groups, it could serve as a valuable scaling law for MSK modeling. In particular, it may aid in estimating the hip joint center using knee width. Nevertheless, further investigation is required, particularly to assess the influence of FSA and femoral anteversion on this relationship.

On the other hand, absolute pelvic measurements only found a significant difference for posterior pelvis width, which was higher in females ($p < 0.05$). This is particularly interesting since anterior pelvis width was found to be similar for both sexes, even when normalized by body height ($p > 0.05$). Given the well-established role of pelvic dimorphism due to reproductive adaptations for females [17], this is unexpected. This observation suggests that relative pelvic width may not be significantly greater in females, contrary to conventional assumptions.

On the other hand, the distance between the left and right anterior superior iliac spine (ASIS) may not be the most accurate measure of pelvis width. As illustrated in Figure 23, the pelvis widens more superiorly, suggesting that a more appropriate measurement may be taken from the iliac crest. This also underscores potential challenges in scaling pelvis geometry, such as pelvis width, based on anatomical landmarks, a common practice in MSK modeling. Relying on such scaling methods may introduce errors in estimating muscle attachment sites and locating hip joint centers, as linear scaling based on bony landmarks has been shown to lead to inaccurate model predictions stemming from these errors [9], [10]. Beyond its implications for MSK modeling, this limitation also restricts our ability to confirm whether true pelvis width differences exist between sexes in our cohort.

Yet, Normalized measurements by body height demonstrated clear sexual dimorphism in the pelvis for our cohort. Pelvic height 1, pelvic height 2, and pelvic depth were all significantly greater in females ($p < 0.001$). While pelvic height was initially expected to be higher in males, this assumption highlights the limitations of defining pelvic height based on anatomical landmarks and joint centers rather than true pelvic morphology. This again questions the validity of linear scaling in capturing such differences.

Pelvic height 1, measured as the distance from the pubic symphysis to the ASIS, appeared greater in females. However, this aligns with the more inferior positioning of the pubic symphysis in females, as observed in Figure 23. Similarly, pelvic height 2, defined as the distance from the hip joint center to the ASIS, was also greater in females. This can be attributed to the higher positioning of the hip joint center in males, along with their deeper pubic arch, which shifts the pubic symphysis more superiorly and posteriorly.

These differences were further validated when pelves were classified based on the sexual dimorphism characteristics mentioned by Fischer et al. (2017) [17] in Figure 23. Furthermore, the male pelvis exhibited less pelvic depth, with the pubic symphysis positioned more superiorly and posteriorly, explaining the lower pelvic height 1 in males. On the other hand, the pubic symphysis for females was located more anteriorly. These

findings align with those of Fischer et al., who demonstrated that males exhibit greater pelvic elongation, whereas females show a more pronounced expansion in the anterior-posterior plane to facilitate childbirth-related adaptations [17].

Conversely, pelvic morphology exhibits substantial individual variability, meaning that not all anatomical differences are strictly sex-dependent. This is consistent with the findings of Delprete et al., who investigated pelvic inlet shape and concluded that sex-based classification alone does not fully account for morphological diversity [42]. Further supporting this, Figure 23 reveals an overlap between some of the male and female pelvis shapes. Interestingly, the "moderate female" shape corresponds more to the "ultra-male" shape, with the sacrum curved inwards, a heightened pelvis, and a narrower inlet shape. The other way around, the "moderate male" shape also shows features of the "ultra-female" shape, with a posterior and outwards pointing sacrum, wider pelvis and rounder inlet shape.

Secondly, muscle volumes retrieved from MRI segmentations were analyzed and compared against values reported in the literature for further validation. Notably, absolute muscle volumes for GMAX and RFEM in this study aligned with previously documented ranges. However, our reported muscle volumes were significantly higher than those found by Lube et al. [36]. This discrepancy is likely due to differences in body height, as our subject group exhibited greater average heights compared to subjects in their study (F:167.3 $\pm$ 4.0 [cm] and M: 174.3 $\pm$ 4.0 [cm]). Given that muscle volume is known to scale with body height [35], this is not an unexpected finding.

On the other hand, the overall muscle volumes reported in this study were lower than those reported by Handsfield et al. [35]. This is assumably caused by the deviation in the partition for both sexes as the Handsfield et al.'s cohort included a higher proportion of male participants, whereas this study had a greater representation of female subjects. Given that males generally exhibit higher absolute muscle volumes, it explains why the overall muscle volume for GMAX and RFEM in the population of their study is higher.

Despite previously reported sex-related differences in muscle mass distribution in the comprehensive overview of Maarleveld et al. (2024), these findings were not fully replicated in our cohort. Although %RFEM was slightly higher in male subjects (F: 1.0% vs. M: 1.2%, p = 0.01), %GMAX was similar between sexes (p > 0.05). This discrepancy may be attributed to differences in normalization methods: in this study, muscle volumes were normalized by lean volume of the lower limbs, whereas Maarleveld et al. normalized muscle volume as a proportion of total muscle volume. This methodological difference likely led to higher proportional muscle volumes in their study, making sex-related discrepancies more pronounced.

Finally, the relationship between bone geometry and muscle volume was examined to determine whether the observed sexual dimorphism in the bones could explain sex-related variations in (proportional) muscle volumes. Several notable positive and negative predictive relationships emerged between

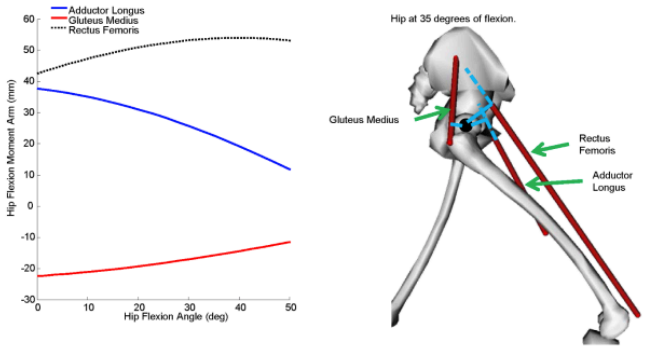


Fig. 10: Hip flexion moment arms of the gluteus medius, adductor longus, and rectus femoris during running [44].

specific bone geometrical measurements and the two muscles investigated, likely reflecting underlying biomechanical factors.

For absolute RFEM volume, no bone metrics were found to be predictive; instead, body weight and sex emerged as the primary predictors, explaining 66.4% of the variance in our cohort (R^2 = 0.664). This is unsurprising, as the analysis considered absolute muscle volume. As demonstrated in our results, male participants exhibited significantly greater RFEM volumes than females (F: 189.2 [cm^3] vs. M: 289.6 [cm^3], p < 0.001). Furthermore, body weight is known to correlate of muscle volume, as it reflects greater muscle mass due to its higher density compared to other soft tissues such as fat, making its predictive role expected in this context [16], [35], [37], [43].

When largely accounting for anthropometric factors such as body weight and sex by using proportional volume, bone metrics remained predictive of %RFEM, albeit to a lesser extent (R^2 = 0.478). Notably, femur length exhibited a negative relationship with %RFEM, which appears counterintuitive given its expected influence on the muscle moment arm for hip flexion. As femur length increases, the RFEM insertion shifts distally, reducing its line-of-action angle with the femur in the sagittal plane (see Figure 11). This shift, combined with femoral sagittal curvature, which is more pronounced in shorter femurs, affects the moment arm. Additionally, the anterior inferior iliac spine (AIIS), the attachment site for RFEM at the pelvis, though not directly measured in this study, likely contributes to this relationship.

Still, a shorter moment arm increases RFEMs force demand to achieve similar joint moments, which should theoretically result in a higher %RFEM in individuals with longer femurs. However, the observed negative relationship with femur len suggests that other anatomical factors, such as muscle insertion variability, may also contribute. Since muscle moment arms for RFEM vary significantly [45], individuals with longer femurs may have a more anteriorly positioned AIIS, altering the moment arm. Given that RFEM achieves its highest moment arm at 35-40 degrees [44], closely aligning with the AIIS (see Figure 10), this landmark appears crucial in determining

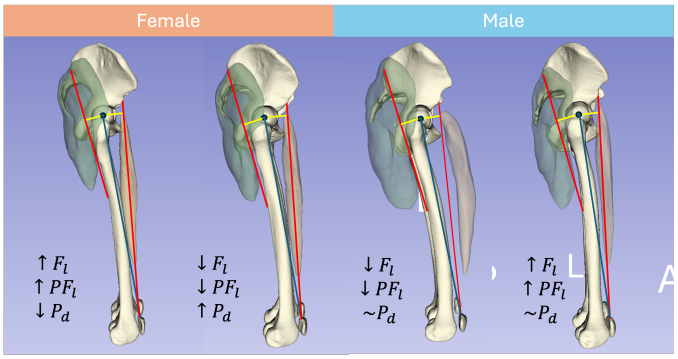


Fig. 11: Moment arms for hip flexion and extension of GMAX and RFEM, based on a subject from our dataset exhibiting relative differences in predictive bone variables. Red lines indicate the line of action for RFEM and GMAX, while yellow lines represent their moment arms. The dot marks the hip joint center.

RFEMs function. Further research is needed to verify this relationship.

Femoral offset appears to influence hip and knee joint biomechanics, as previously mentioned [40], and may, therefore, be predictive of %RFEM. Given RFEMs role in both hip flexion and knee extension during gait [44], [46], alterations in hip and knee joint force patterns due to higher femoral offset may increase RFEMs compensatory force demand. Although RFEM does not directly counteract lateral or medial joint reaction forces, a study found that its activation increases when hip abductor strength, particularly in the gluteus medius and minimus, is reduced due to higher femoral offset [47]. Reduced strength in the abductors may result from greater femoral offset, as their moment arms increase and therefore reduced force demand to create similar joint moment [48]. Consequently, muscle recruitment patterns may differ for those with higher femoral offset, with RFEM compensating more for joint stability. This could explain why males exhibited higher %RFEM, as their femoral offset was significantly greater than that of females ($p < 0.001$). Conversely, sexual dimorphism in femoral offset may also explain the smaller partition of the gluteus medius in muscle mass distribution for males [13]. Alternatively, femoral offset may simply reflect the width of the Rectus Femoris, serving as an indirect indicator of its muscle volume. As shown in Appendix K, males appear to have wider RFEMs.

Similar to RFEM, absolute GMAX volume was primarily influenced by sex (R^2 = 0.679). However, bone metrics such as pelvis depth and femur length also contributed to absolute GMAX volume. Femur length likely affects the moment arm for hip extension, where GMAX serves as the primary contributor [49]. As femur length increases, GMAX attaches more distally on the femur, shifting its line of action farther from the hip joint center and increasing its moment arm (see Figure 11). This shift would reduce the force demand on GMAX for generating joint moments. However, similar to RFEM, GMAX exhibits considerable variation in its femoral insertion site across individuals [45], leading to variability in moment arms and making this effect more subject-specific. As shown in Figure 11, females with longer femurs appear to have a more proximal femoral attachment compared to those with shorter femurs.

Additionally, %GMAX was negatively influenced by the pelvis-femur length metric, which is largely determined by femur length. However, this metric predicted %GMAX only to a limited extent (R^2 = 0.151). This further suggests that femur length may contribute to an increased moment arm of GMAX for hip flexion and could, therefore, be a potential predictor of its proportional volume. As females in this study had relatively longer femurs than males (p = 0.01), the regression equations suggest this would result in lower %GMAX and absolute GMAX volumes. This contrasts with the findings of Maarleveld et al. [13], who reported that %GMAX was 17% higher in females compared to males. However, this difference was not confirmed in the present study. One possible explanation is the relatively tall female cohort in this study. Additionally, males typically exhibit greater femur length due to increased body height [37], which could contribute to the lower %GMAX observed in males by Maarleveld et al.

The positive predictive relationship of pelvis depth with GMAX volume appears contradictory. A deeper pelvis could shift the posterior superior iliac spine further posteriorly, thereby repositioning GMAX attachment points. This would increase the angle between the muscles line of action for hip flexion and the femoral shaft, resulting in a larger moment arm and consequently reducing the force demand on GMAX. Likewise, it seems to increase the moment arm for exorotation due to this more posteriorly positioned PSIS (see Figure 25 in Appendix M). Therefore, its positive association with GMAX volume seems unexpected. However, since this relationship pertains to absolute GMAX volume, pelvis depth may primarily reflect the width of GMAX rather than its force-generating capacity. From a superior view, pelvis depth closely relates to the width span of GMAX, suggesting that a higher width relates to greater muscle volume.

Taking this all together, using proportional muscle volumes allows for a more precise analysis of the biomechanical implications of skeletal dimorphism, as absolute muscle volumes remain influenced by anthropometric factors such as sex and body weight. Interestingly, body height was not a determining factor for absolute muscle volume, despite its previously established strong relationship with muscle volume [35]. While proportional volumes provide insight into sexual dimorphism, the regression models explained less variance (%GMAX: $R^2 = 0.151$, %RFEM: $R^2 = 0.478$) compared to those for absolute muscle volumes (GMAX: $R^2 = 0.679$, RFEM: $R^2 = 0.664$). This was particularly evident for %GMAX, which had a lower R^2, likely due to having only one significant independent variable. Given that %GMAX and %RFEM may better capture the biomechanical effects of sexual dimorphism, it is notable that their predictive models included only femoral or femur-to-pelvis bone metrics, suggesting a stronger reliance

on femoral sexual dimorphism. Nevertheless, conclusions regarding the influence of sexual dimorphism on %GMAX remain inconclusive, as no significant sex-related differences were observed.

V. Limitations & Recommendations

This study has several limitations that should be addressed in future research. Firstly, the measurement of bone metrics was constrained, as it relied solely on marker positions used in MSK modeling, along with joint centers and bone segment centers, rather than direct anatomical measurements. This limitation may have affected the accuracy of the extracted bone geometry. To improve precision, incorporating additional anatomical landmarks, such as the proximal point of the ilium or the distal point of the pubic arch, could provide better estimates of bone metrics (e.g., pelvic height and width) and enhance assessments of sexual dimorphism in our cohort. This refinement could also further impact the analysis of relationships with proportional muscle volumes, potentially revealing new and meaningful associations through these additional measurements, as this study has already revealed. Secondly, the use of lean volume to derive proportional muscle volumes posed a limitation. This approach resulted in significantly lower proportional muscle volumes compared to those reported by Maarleveld et al., suggesting that lean volume does not predominantly reflect muscle volume. Similarly, using fat percentage derived through BIA may lack the precision needed to accurately estimate fat volume. Additionally, assuming uniform fat distribution could have further impacted accuracy. Consequently, we were unable to confirm sex-related differences in lower-extremity muscle volume distribution for GMAX. For future research, segmenting all lower-extremity muscles in this cohort would provide more accurate proportional estimates. It would also allow for the exploration of muscle leverage interdependencies through these bone metrics. For instance, as femoral offset increases, the force demand and, consequently, the strength in the hip abductors decrease. At the same time, the proportion of RFEM therefore might increase. Further investigation into these relationships could provide deeper biomechanical insights. Thirdly, this study included only sixteen young adults, limiting the generalizability of the findings to other age groups and larger populations. Therefore, future research should investigate if these relationships persist in large cohorts. Lastly, this study did not acquire data on muscle insertion points at the femur and pelvis. This limitation restricted our ability to interpret moment arms as they are highly dependent on muscle attachment sites. Future research should incorporate muscle attachment points to draw more accurate conclusions about the implications of sexual dimorphism on moment arms. This would help determine whether sex-specific variations in moment arms can explain the observed sex-related differences in proportional muscle volumes.

VI. Conclusion

This study aimed to investigate whether sexual dimorphism in the femur and pelvis could explain sex-related differences in the (proportional) muscle volumes of the Gluteus Maximus (GMAX) and Rectus Femoris (RFEM). To achieve this, an automatic segmentation model was developed to extract the lower-extremity skeleton and the two muscles of interest from MRI scans of sixteen young adults. The model outperformed existing MRI-based segmentation methods, demonstrating its effectiveness.

A. Key Findings

1) Sexual Dimorphism in Bone Geometry:

- Femoral offset and knee width were larger in males, even after normalization by body height. However, contrary to conventional assumptions, femur length showed no significant sex difference but was proportionally greater in females when normalized.
- Pelvic depth and posterior pelvic width were greater in females, whereas common pelvic metrics such as pelvic height and traditionally measured pelvic width did not show significant sex differences. This raises questions about whether the measurement methods accurately capture these anatomical features.

2) Sex-Related Differences in Proportional Muscle Volumes:

- Absolute volumes for RFEM and GMAX were higher in males, as expected, and aligned with previous literature.
- %RFEM was significantly higher in males, whereas %GMAX did not differ significantly between sexes.

3) Relationship Between Bone Geometry and Muscle Volume:

- Femoral offset was partially associated with sex-related differences in %RFEM. Since femoral offset varies between sexes, it may contribute to differences in hip and knee joint reaction forces, muscle leverage, and RFEM's role in joint stability, particularly during gait.
- While sexual dimorphism in the femur and pelvis may influence %GMAX, the lack of a significant sex difference in %GMAX prevents definitive conclusions.

B. Final Implications

These findings suggest that sex-based differences in femoral offset may contribute to variations in muscle function and joint biomechanics, particularly in RFEMs role in knee and hip stability. However, further research is needed to fully understand the biomechanical implications of sexual dimorphism in the lower-extremity skeleton.

References

[1] M. J. Legato, *Eve's rib: The New Science of Gender-Specific Medicine and How It Can Save Your Life*. Crown Publications, 9 2014.

[2] S. T. d. Vries, P. Denig, C. Ekhart, J. Burgers, N. Kleefstra, P. G. M. Mol, and E. v. Puijenbroek, "Sex differences in adverse drug reactions reported to the national pharmacovigilance centre in the netherlands: an explorative observational study," *British Journal of Clinical Pharmacology*, vol. 85, pp. 1507–1515, 2019.

[3] L. L. Tosi, "Does Sex Matter in Musculoskeletal Health? The Influence of Sex and Gender on Musculoskeletal Health," *Journal of Bone and Joint Surgery*, vol. 87, no. 7, p. 1631, 7 2005. [Online]. Available: https://doi.org/10.2106/jbjs.e.00218

[4] A. K. Ramachandran, J. S. Pedley, S. Moeskops, J. L. Oliver, G. D. Myer, and R. S. Lloyd, "Changes in lower limb biomechanics across various stages of maturation and implications for acl injury risk in female athletes: a systematic review," *Sports Medicine*, vol. 54, pp. 1851–1876, 2024.

[5] T. A. Donelon, J. Edwards, M. Brown, P. A. Jones, J. ODriscoll, and T. DosSantos, "Differences in Biomechanical Determinants of ACL injury Risk in Change of Direction Tasks between Males and Females: A Systematic Review and Meta-Analysis," *Sports Medicine - Open*, vol. 10, no. 1, 4 2024. [Online]. Available: https://doi.org/10.1186/s40798-024-00701-z

[6] E. van der Kruk, "Biasmechanics: Does an unconscious bias still persist in biomechanics, positioning males as the default in human research? a meta-analysis on the journal of biomechanics 2024 publications," *Journal of Biomechanics*, vol. 181, p. 112560, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021929025000715

[7] S. Delp, J. Loan, M. Gross, F. Zajac, E. Topp, and J. Rosen, "An interactive graphics-based model of the lower extremity to study orthopaedic surgical procedures," *IEEE transactions on bio-medical engineering*, vol. 37, pp. 757–67, 09 1990.

[8] B. Bolsterlee, D. Veeger, and E. Chadwick, "Clinical applications of musculoskeletal modelling for the shoulder and upper limb," *Medical biological engineering computing*, vol. 51, 07 2013.

[9] L. Scheys, A. Spaepen, P. Suetens, and I. Jonkers, "Calculated moment-arm and muscle-tendon lengths during gait differ substantially using mr based versus rescaled generic lower-limb musculoskeletal models," *Gait Amp; Posture*, vol. 28, pp. 640–648, 2008.

[10] R. Akhundov, D. J. Saxby, L. E. Diamond, S. Edwards, P. Clausen, K. Dooley, S. Blyton, and S. J. Snodgrass, "Is subject-specific musculoskeletal modelling worth the extra effort or is generic modelling worth the shortcut?" *Plos One*, vol. 17, p. e0262936, 2022.

[11] A. Rajagopal, C. L. Dembia, M. S. DeMers, D. D. Delp, J. L. Hicks, and S. L. Delp, "Full-body musculoskeletal model for muscle-driven simulation of human gait," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 10, pp. 2068–2079, 2016.

[12] V. Carbone, R. Fluit, P. Pellikaan, M. M. v. d. Krogt, D. Janssen, M. Damsgaard, L. Vigneron, T. Feilkas, B. Koopman, and N. Verdonschot, "Tlem 2.0 a comprehensive musculoskeletal geometry dataset for subject-specific modeling of lower extremity," *Journal of Biomechanics*, vol. 48, pp. 734–741, 2015.

[13] R. Maarleveld, H. E. J. Veeger, V. D. H. F. C. T, J. Son, R. L. Lieber, and V. D. K. E, "What the %pcsa? addressing diversity in lower-limb musculoskeletal models: age- and sex-related differences in pcsa and muscle mass," *arXiv (Cornell University)*, 10 2024. [Online]. Available: http://arxiv.org/abs/2411.00071

[14] V. Carbone, v. d. M. Krogt, B. Koopman, and N. Verdonschot, "Sensitivity of subject-specific models to hill muscletendon model parameters in simulations of gait," *Journal of Biomechanics*, vol. 49, pp. 1953–1960, 2016.

[15] J. R. Meakin, J. Fulford, R. Seymour, J. R. Welsman, and K. M. Knapp, "The relationship between sagittal curvature and extensor muscle volume in the lumbar spine," *Journal of Anatomy*, vol. 222, pp. 608–614, 6 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/joa.12047

[16] B. Preininger, K. Schmorl, P. v. Roth, T. Winkler, G. Matziolis, C. Perka, and S. Tohtz, "A formula to predict patients gluteus medius muscle volume from hip joint geometry," *Manual Therapy*, vol. 16, pp. 447–451, 2011.

[17] B. Fischer and P. Mitteroecker, "Allometry and sexual dimorphism in the human pelvis," *The Anatomical Record*, vol. 300, no. 4, pp. 698–705, 2017. [Online]. Available: https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/ar.23549

[18] L. Modenese and J.-B. Renault, "Automatic generation of personalized skeletal models of the lower limb from three-dimensional bone geometries," *Journal of Biomechanics*, vol. 116, p. 110186, 2021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929020306102

[19] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[20] M. Brett, C. J. Markiewicz, M. Hanke, M. Cottaar, F. C. Morency, Y. O. Halchenko, D. Wassermann, S. Gerhard, and S. S. Ghosh, "NiBabel: Processing and analyzing neuroimaging data in Python," *Frontiers in Neuroinformatics*, vol. 17, p. 48, 2023.

[21] A. Fedorov, R. Beichel, J. KalpathyCramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. M. Fennessy, M. Sonka, J. M. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging*, vol. 30, pp. 1323–1341, 2012.

[22] T. A. D'Antonoli, L. K. Berger, A. K. Indrakanti, N. Vishwanathan, J. WeiSS, M. Jung, Z. Berkarda, A. Rau, M. Reisert, T. Küstner, A. Walter, E. M. Merkle, M. Segeroth, J. Cyriac, S. Yang, and J. Wasserthal, "Totalsegmentator mri: Sequence-independent segmentation of 59 anatomical structures in mr images," 2024. [Online]. Available: https://arxiv.org/abs/2405.19492

[23] C. van Straaten, "Automated Localisation of Subject-Specific Muscle-Tendon Paths of the Lower Limbs Using nnU-Net on Magnetic Resonance Images," Master's thesis, Delft University of Technology, January 2025. [Online]. Available: https://repository.tudelft.nl/record/uuid:c7694346-3fb3-4427-b571-1046859616?4

[24] W. H. Henson, C. Mazzá, and E. DallAra, "Augmented lower limb MR images," 12 2022. [Online]. Available: https://orda.shef.ac.uk/articles/dataset/Augmented_lower_limb_MR_images/20440164

[25] ——, "Augmented images associated segmentations," 12 2022. [Online]. Available: https://orda.shef.ac.uk/articles/dataset/Augmented_images_associated_segmentations/20440203

[26] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. KoppSchneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, B. v. Ginneken, M. Bilello, P. Bilic, P. F. Christ, K. G. Richard, M. J. Gollub, S. Heckers, H. Huisman, W. R. Jarnagin, M. McHugo, S. Napel, J. S. G. Pernicka, K. Rhode, C. Tobon-Gómez, E. Vorontsov, J. Meakin, S. Ourselin, M. Wiesenfarth, P. Arbeláez, B. Bae, S. Chen, L. Daza, J. Feng, B. He, F. Isensee, Y. Ji, F. Jia, N. Kim, I. Kim, D. Merhof, A. Pai, B. Park, M. Perslev, R. Rezaiifar, O. Rippel, I. Sarasúa, W. Shen, J. Son, C. Wachinger, L. Wang, Y. Wang, Y. Xia, D. Xu, Z. Xu, Y. Zheng, A. L. Simpson, L. Maier-Hein, and M. J. Cardoso, "The medical segmentation decathlon," *Nature Communications*, vol. 13, 2022.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science*, pp. 234–241, 2015.

[28] F. Isensee, T. Wald, C. Ulrich, M. Baumgartner, S. Roy, K. Maier-Hein, and P. F. Jaeger, "nnu-net revisited: A call for rigorous validation in 3d medical image segmentation," *arXiv (Cornell University)*, 4 2024. [Online]. Available: http://arxiv.org/abs/2404.09556

[29] Delft High Performance Computing Centre (DHPC), "DelftBlue Supercomputer (Phase 2)," https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2, 2024.

[30] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001–. [Online]. Available: http://www.scipy.org/

[31] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference*, V. Scarano, R. D. Chiara, and U. Erra, Eds. The Eurographics Association, 2008.

[32] T. M. Inc., "Matlab version: R2024b," Natick, Massachusetts, United States, 2024. [Online]. Available: https://www.mathworks.com

[33] S. Heymsfield, *Human body composition*. Human kinetics, 2005, vol. 918.

[34] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Online]. Available: https://doi.org/10.21105/joss.03021

[35] G. G. Handsfield, C. H. Meyer, J. M. Hart, M. F. Abel, and S. S. Blemker, "Relationships of 35 lower limb muscles to height and body mass quantified using mri," *Journal of Biomechanics*, vol. 47, pp. 631–638, 2014.

[36] J. Lube, S. Cotofana, I. Bechmann, T. L. Milani, O. Özkurtul, T. Sakai, H. Steinke, and N. Hammer, "Reference data on muscle volumes of healthy human pelvis and lower extremity muscles: an in vivo magnetic resonance imaging feasibility study," *Surgical and Radiologic Anatomy*, vol. 38, no. 1, p. 97106, aug. 2015.

[37] J. Son, S. R. Ward, and R. L. Lieber, "Scaling relationships between human leg muscle architectural properties and body size," *Journal of Experimental Biology*, vol. 227, no. 6, mar. 2024.

[38] M. Frysz, J. Gregory, R. M. Aspden, L. Paternoster, and J. H. Tobias, "Sex differences in proximal femur shape: findings from a population-based study in adolescents," *Scientific Reports*, vol. 10, no. 1, 3 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-61653-4

[39] S. P. Tuck, M. S. Pearce, D. J. Rawlings, F. N. Birrell, L. Parker, and R. M. Francis, "Differences in bone mineral density and geometry in men and women: the newcastle thousand families study at 50 years old," *British Journal of Radiology*, vol. 78, no. 930, pp. 493–498, 02 2014. [Online]. Available: https://doi.org/10.1259/bjr/42380498

[40] H. A. Rüdiger, M. Guillemin, A. Latypova, and A. Terrier, "Effect of changes of femoral offset on abductor and joint reaction forces in total hip arthroplasty," *Archives of Orthopaedic and Trauma Surgery*, vol. 137, pp. 1579–1585, 2017.

[41] M. Scorcelletti, N. D. Reeves, J. Rittweger, and A. Ireland, "Femoral anteversion: significance and measurement," *Journal of Anatomy*, vol. 237, no. 5, pp. 811–826, 6 2020. [Online]. Available: https://doi.org/10.1111/joa.13249

[42] H. Delprete, "Pelvic inlet shape is not as dimorphic as previously suggested," *The Anatomical Record*, vol. 300, no. 4, pp. 706–715, mar. 2017. [Online]. Available: https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/ar.23544

[43] B. Chen, T. Shih, C. Hsu, C. Yu, S. Wei, C. Chen, C. H. Wu, and C.-Y. Chen, "Thigh muscle volume predicted by anthropometric measurements and correlated with physical function in the older adults," *The Journal of nutrition, health and aging*, vol. 15, pp. 433–438, 6 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1279770723006292

[44] A. Schmitz, J. Norberg, K. Snarski, and D. Piovesan, "Association Between Impact Peak and Hip Flexor Activity During Running," 11 2016. [Online]. Available: https://doi.org/10.1115/imece2016-65374

[45] G. N. Duda, D. Brand, S. Freitag, W. Lierse, and E. Schneider, "Variability of femoral muscle attachments," *Journal of Biomechanics*, vol. 29, no. 9, pp. 1185–1190, 9 1996. [Online]. Available: https://doi.org/10.1016/0021-9290(96)00025-5

[46] S. J. Piazza and S. L. Delp, "The influence of muscles on knee flexion during the swing phase of gait," *Journal of Biomechanics*, vol. 29, no. 6, pp. 723–733, 6 1996. [Online]. Available: https://doi.org/10.1016/0021-9290(95)00144-1

[47] M. Güngörürler and H. Havtçolu, "Effects of weak gluteal muscles and increased femoral offset on muscle activations and stresses on femur after total hip arthroplasty," *Orthopaedic Proceedings*, vol. 100-B, no. SUPP_16, pp. 67–67, 2018. [Online]. Available: https://doi.org/10.1302/1358-992X.2018.16.067

[48] H. A. Rüdiger, V. Parvex, and A. Terrier, "Impact of the femoral head position on moment arms in total hip arthroplasty: a parametric finite element study," *The Journal of Arthroplasty*, vol. 31, pp. 715–720, 2016.

[49] O. Ludwig, C. Dindorf, S. Kelm, J. Kelm, and M. Fröhlich, "Muscular strategies for correcting the pelvic position to improve posturean exploratory study," *Journal of Functional Morphology and Kinesiology*, vol. 9, p. 25, 2024.

APPENDIX A
NOMENCLATURE

α_1	*Pelvis angle 1*
α_2	*Pelvis angle 2*
ASIS	*anterior superior iliac spine*
AIIS	*anterior inferior iliac spine*
CT	*Computed tomography*
DSC	*DICE Similarity Coefficient*
F_l	*Femur length*
FSA	*Femoral neck shaft angle*
F_o	*Femoral offset*
GMAX	*Gluteus Maximus (volume)*
%GMAX	*proportional Gluteus Maximus volume*
HD	*Hausdorff Distance*
MSK	*Musculoskeletal*
MRI	*Magnetic resonance imaging*
MV	*Muscle volume*
P_d	*Pelvis depth*
P_{h1}	*Pelvis height 1*
P_{h2}	*Pelvis height 2*
P_w^a	*Pelvis width anterior*
P_w^p	*Pelvis width posterior*
PSIS	*posterior superior iliac spine*
RFEM	*Rectus Femoris (volume)*
%RFEM	*proportional Rectus Femoris volume*

APPENDIX B
SUBJECT DATA

Subject ID	**Sex**	**Age**	**Height** [cm]	**Weight** [kg]	**Body fat percentage** [$\%$]
101	M	34	185.0	66.2	8.7
002	F	29	171.0	64.5	34.3
003	F	26	174.0	57.9	22.5
004	F	27	180.0	68.3	33.3
005	F	30	169.5	72.5	36.2
006	F	31	165.5	55.5	26.6
107	M	31	172.0	65.1	16.3
108	M	24	162.5	68.5	20.6
009	F	28	169.3	65.7	30.4
112	M	28	183.5	80.0	17.0
013	F	24	172.0	63.4	25.4
114	M	27	195.5	72.9	12.6
115	M	30	182.3	71.0	13.5
016	F	26	174.5	68.2	30.0
017	F	32	176.0	57.8	24.3
118	M	36	185.0	94.0	23.4

TABLE VI: Antropometric data for all the subjects included in this study

APPENDIX C
OVERVIEW OF SEGMENTED BONES

Bone	***# per subject***
Sacrum	1
Pelvis	2
Femur	2
Patella	2
Tibia	2
Fibula	2
Calcaneus	2
Talus	2
Navicular	2
Foot bones	2
Total	19

TABLE VII: Overview of all bones segmented in this study

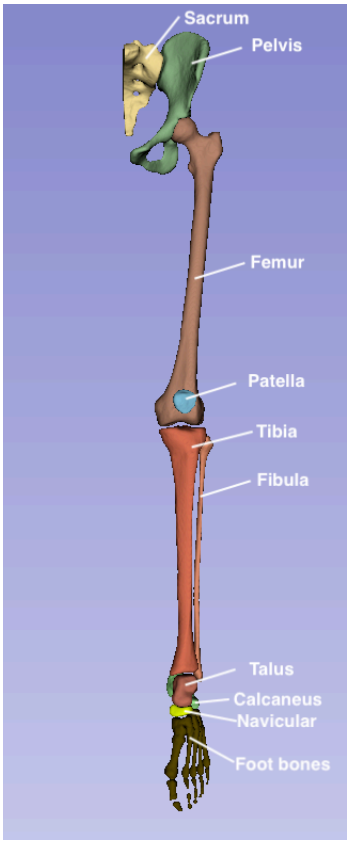


Fig. 12: A sample from the training set for nnU-Net, each color represent different segment

APPENDIX D
EXCLUDED BONE GEOMETRICAL MEASUREMENTS

Variable	Measurement	Reason for exclusion
Bodyweight leverarm (BLA)	euclidian distance between pubic symphisus and hip joint center	bilateral asymmetry
Hip curvature (Hip_Curv)	euclidian distance between greater trochanter of the femur and PSIS	inconclusive results for sexual dimorphism and no predictive value for muscle volume
Hip Triangulation (Hip_Tri)	euclidian distance between hip joint center and PSIS	no predictive value for muscle volume
Pelvis angle 1 (α_1)	the angle between ASIS and PSIS from the transverse plane	accuracy of metric debatable and no predictive value for muscle volume
Pelvis angle 2 (α_2)	angle between the center of the pelvis and the ASIS	accuracy of metric debatable and no predictive value for muscle volume
Pelvis tilt	rotation of pelvis by STAPLE in the saggittal plane	accuracy of metric debatable and no predictive value for muscle volume

TABLE VIII: Excluded bone metrics listed together with the reason for exclusion in the results.

APPENDIX E
OVERVIEW AUTOMATIC PIPELINE FOR DERIVATION OF BONE GEOMETRY AND MUSCLE VOLUME

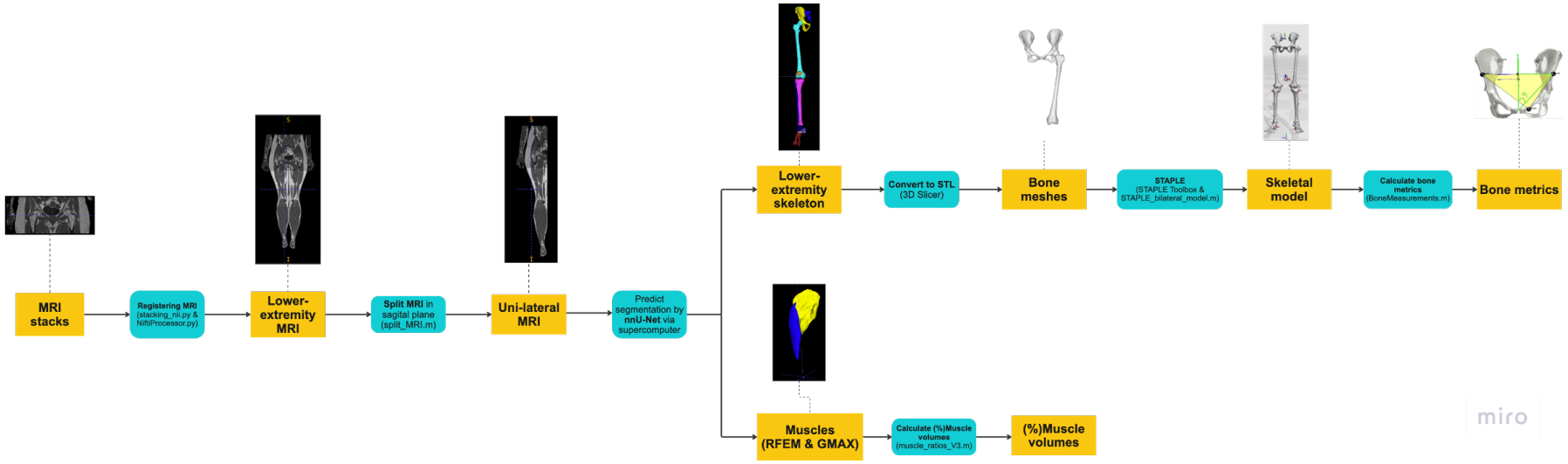


Fig. 13: Flowscheme for the automatic pipeline to derive the bone metrics and muscle volumes from MRI data. Between bracks, scripts or software are mentioned that were utilized to achieve certain steps. Blue blocks represent input/output and orange blocks represent the operation.

APPENDIX F
T-TESTS FOR BILATERAL SYMMETRY ASSESSMENT

Bone metrics

Variable	Test	Statistic	p-value
Bodyweight Leverarm	Paired t-test	-3.568	0.003
Femoral Offset	Paired t-test	-2.854	0.182
Femoral Offset / Knee Width	Paired t-test	-1.954	0.100
Femur Length	Paired t-test	1.328	0.204
Hip Curvature	Paired t-test	-1.995	0.065
Hip Triangulation	Paired t-test	0.710	0.489
Knee Width	Paired t-test	-0.505	0.621
Pelvic-Femur Length	Paired t-test	0.682	0.506
Pelvis Angle 1	Paired t-test	-2.273	0.038
Pelvis Angle 2	Wilcoxon test	43.000	0.211
Pelvis Depth	Paired t-test	-1.122	0.279
Pelvis Height 1	Wilcoxon test	48.000	0.323
Pelvis Height 2	Paired t-test	0.223	0.827

Muscle volumes

Variable	Test	Statistic	p-value
GMAX	Paired t-test	-1.127	0.278
%GMAX	Paired t-test	-1.122	0.280
RFEM	Paired t-test	-0.994	0.336
%RFEM	Paired t-test	-0.830	0.420

TABLE IX: T-test results for assessing bilateral symmetry. If $p > 0.05$, variables were significantly considered bilateral symmetrical. If Levenes test indicated homogeneity of variance ($p > 0.05$), a Paired t-test was applied; otherwise, a Wilcoxon t-test was used.

APPENDIX G
T-TEST RESULTS FOR EVALUATION OF SEX-RELATED DIFFERENCES

Anthropometrics

Variable	Levene p-value	T-test p-value	Mean (Male)	Mean (Female)	Higher in
BH $[cm]$	0.063	0.003	180.829	172.422	Male
BW $[kg]$	0.251	0.001	73.957	63.756	Male

Bone metrics

Variable	Levene p-value	T-test p-value	Mean (Male)	Mean (Female)	Higher in
Bodyweight leverarm $[cm]$	0.774	0.221	10.282	10.471	Female
Femur length $[cm]$	0.061	0.286	43.037	42.074	Male
Femoral offset $[cm]$	0.203	0.002	7.331	6.741	Male
Femoral offset / Knee width	0.867	0.784	0.876	0.881	Female
Hip Curvature $[cm]$	0.186	0.059	17.709	17.052	Male
Hip Triangulation $[cm]$	0.713	0.946	13.317	13.328	Female
Knee Width $[cm]$	0.148	<0.001	8.361	7.659	Male
Pelvic Tilt [deg]	0.963	0.429	-11.098	-12.882	Female
Pelvis Angle 1* [deg]	0.010	0.033	20.029	22.880	Female
Pelvis Angle 2 [deg]	0.785	0.452	89.253	89.396	Female
Pelvis Depth $[cm]$	0.519	0.427	15.454	15.672	Female
Pelvis Height 1 $[cm]$	0.296	0.483	14.860	15.093	Female
Pelvis Height 2 $[cm]$	0.626	0.630	9.601	9.699	Female
Pelvis Width Anetrior $[cm]$	0.471	0.449	23.042	22.247	Male
Pelvis Width Posterior $[cm]$	0.444	0.046	9.061	10.362	Female
Pelvis-Femur length $[cm]$	0.089	0.298	50.883	49.854	Male

Muscle volumes

Variable	Levene p-value	T-test p-value	Mean (Male)	Mean (Female)	Higher in
RFEM $[cm^3]$	0.246	0.000	285.985	194.155	Male
GMAX $[cm^3]$	0.196	0.000	990.594	756.673	Male
RFEM [%]	0.711	0.011	1.192	1.049	Male
GMAX [%]	0.646	0.727	4.153	4.086	Male

TABLE X: T-test results for testing sex-related differences for all metrics, being anthropometrics, bone metrics and muscle volumes. The Levenes test p-value evaluates the homogeneity of variance for each variable. If p > 0.05 for the Levene's test, equal variance was assumed and a Student's test was chosen over a Welch's t-test (see * for these variables).

Bone metrics

Variable	Levene p-value	T-test p-value	Mean (Male)	Mean (Female)	Higher in
Bodyweight leverarm	0.833	<0.001	0.057	0.061	Female
Femur length	0.984	0.013	0.238	0.243	Female
Femoral offset	0.273	0.035	0.040	0.039	Male
Hip Curvature	0.250	0.535	0.098	0.099	Female
Hip Triangulation	0.767	0.002	0.074	0.077	Female
Knee Width	0.052	0.014	0.046	0.044	Male
Pelvis Depth*	0.009	<0.001	0.085	0.090	Female
Pelvis Height 1	0.602	<0.001	0.082	0.088	Female
Pelvis Height 2	0.087	<0.001	0.053	0.056	Female
Pelvis width anterior	0.174	0.725	0.127	0.129	Female
Pelvis width posterior	0.622	0.017	0.050	0.060	Female
Pelvis-femur length	0.651	0.003	0.281	0.289	Female

Muscle volumes

Variable	Levene p-value	T-test p-value	Mean (Male)	Mean (Female)	Higher in
RFEM $[cm^2]$	0.479	<0.001	1.585	1.128	Male
GMAX $[cm^2]$	0.138	<0.001	5.484	4.396	Male

TABLE XI: T-test results assessing sex-related differences in bone metrics and muscle volumes normalized by body height. Angular metrics and ratios, being independent of height, were excluded from the analysis. If Levenes test indicated homogeneity of variance (p > 0.05), a Students t-test was applied; otherwise, a Welchs t-test was used (* indicates these variables).

APPENDIX H
CROSS-SECTIONAL RESULTS FOR BONE AND MUSCLE SEGMENTATION MODELS

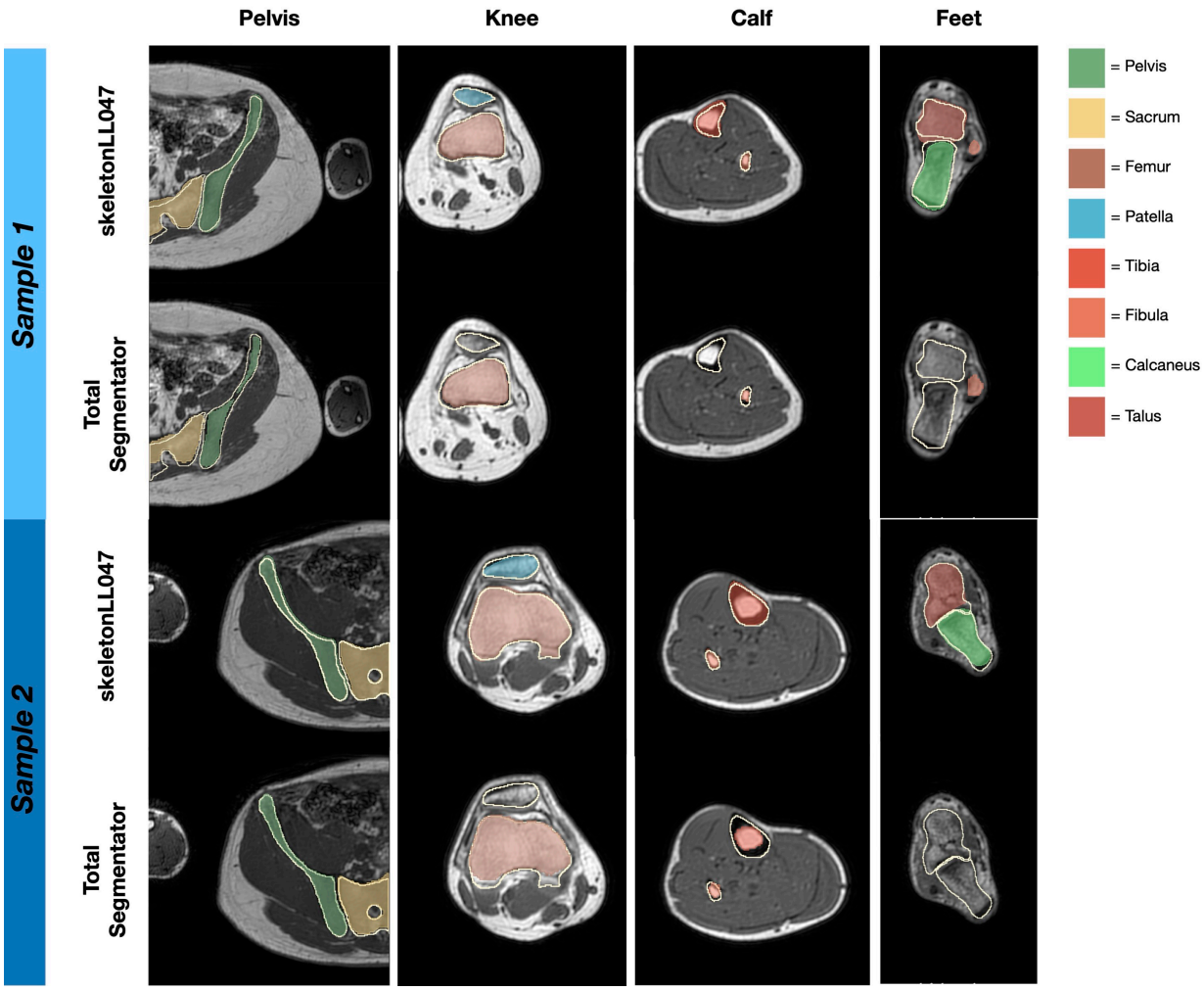


Fig. 14: Results for the automatic bone segmentation models illustrated in the transverse plane across different levels across the leg. The white outlines represent the ground truth, whereas the transparent overlays visualize the predictions by the models.

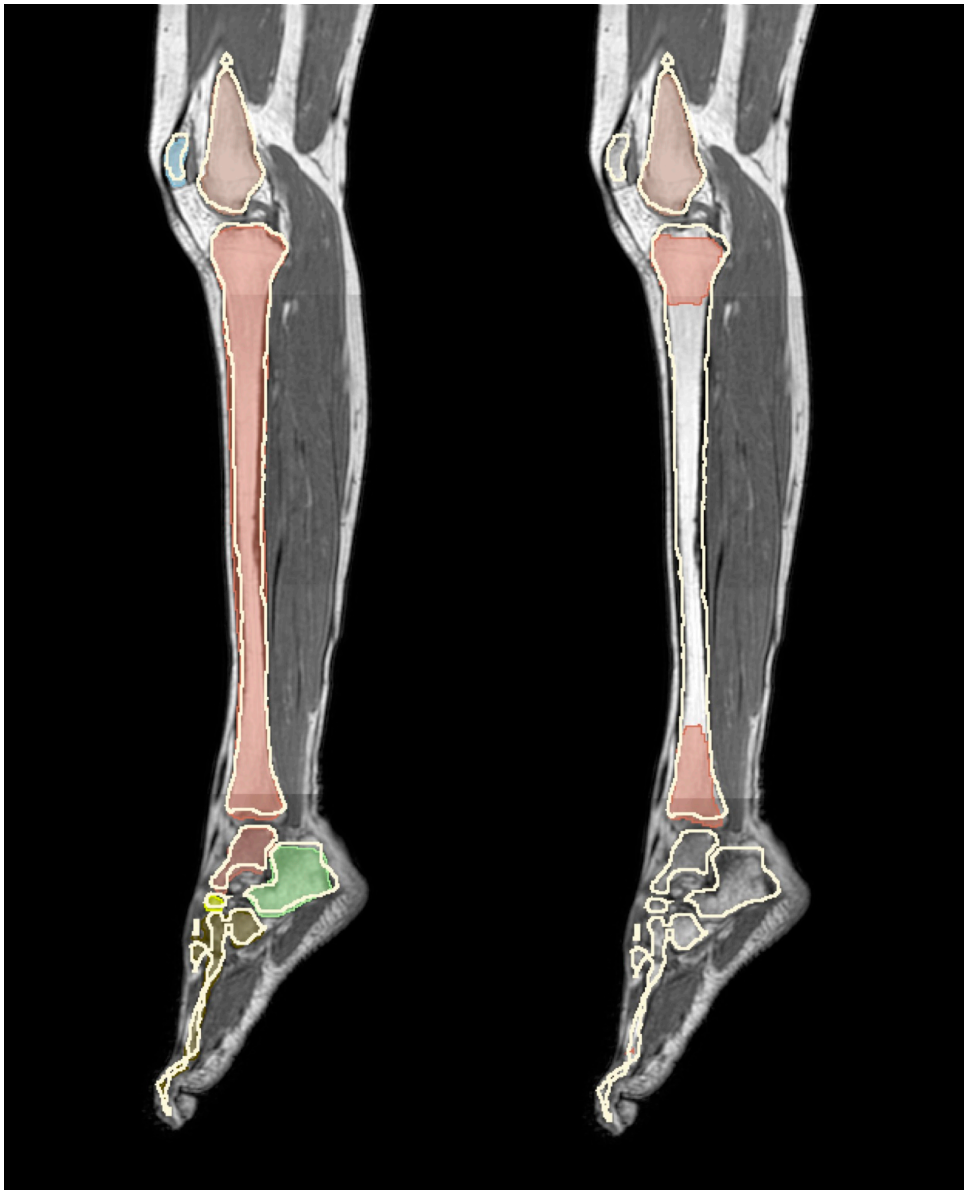


Fig. 15: Predictions of the lower leg by the bone segmentation models for sample 1 in the sagittal plane. Left: skeletonLL047, right: TotalSegmentator, white outlines represent the ground truth.

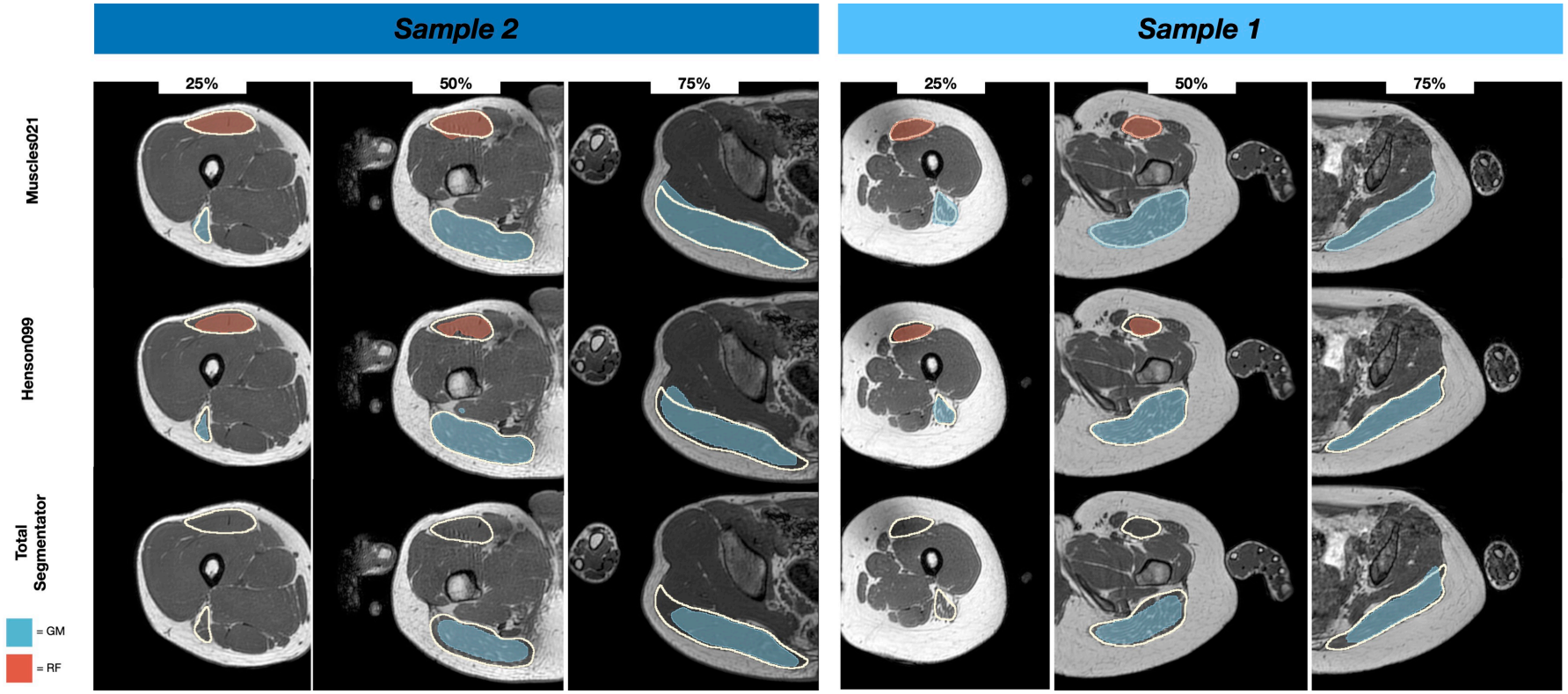


Fig. 16: Results for different automatic muscle segmentation models illustrated in the transverse plane across three levels across the upper leg. The white outlines represent the ground truth, whereas the transparent overlays visualize the predictions by the models.

APPENDIX I
TRAINING SET MODELS

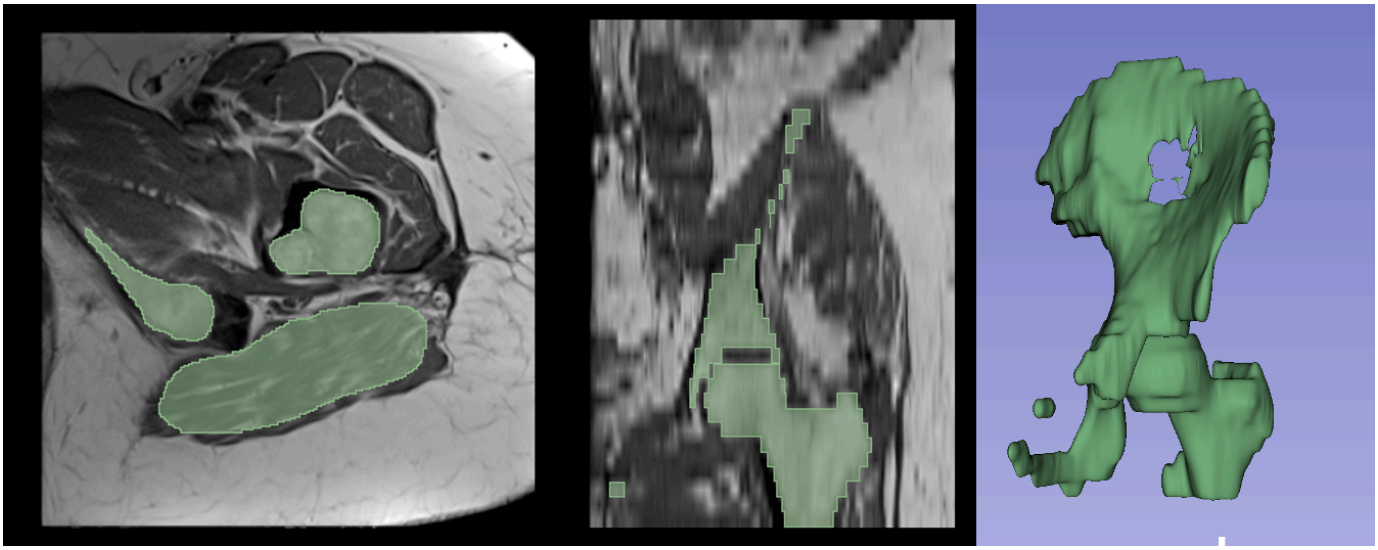


Fig. 17: A sample of the TotalSegmentator training set, illustrating (lacking) segmentations around the hip joint

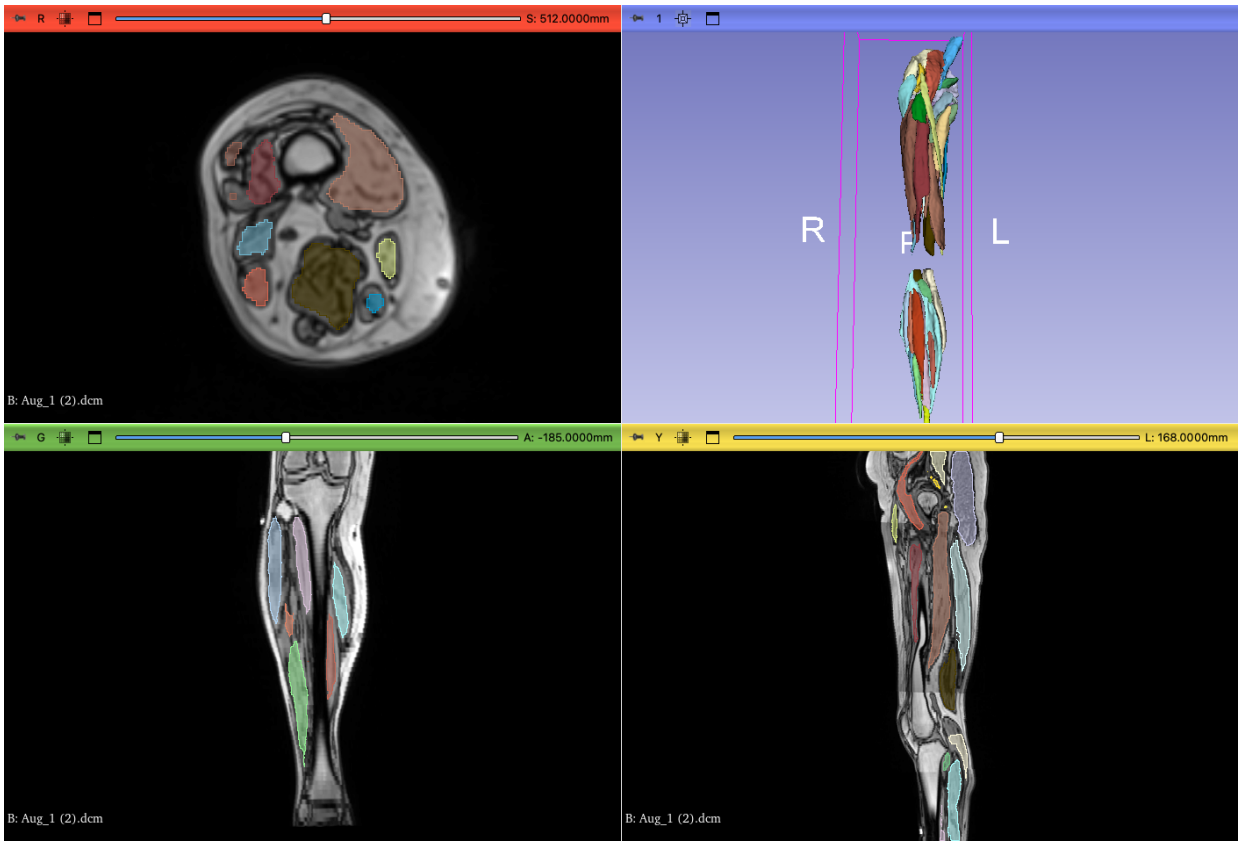


Fig. 18: A sample of the Henson099 training set, displaying the lacking segmentation of some of 35 muscles in the lower extremity

APPENDIX J
SEGMENTATION MODEL TESTED ON OTHER MRI MODALITY

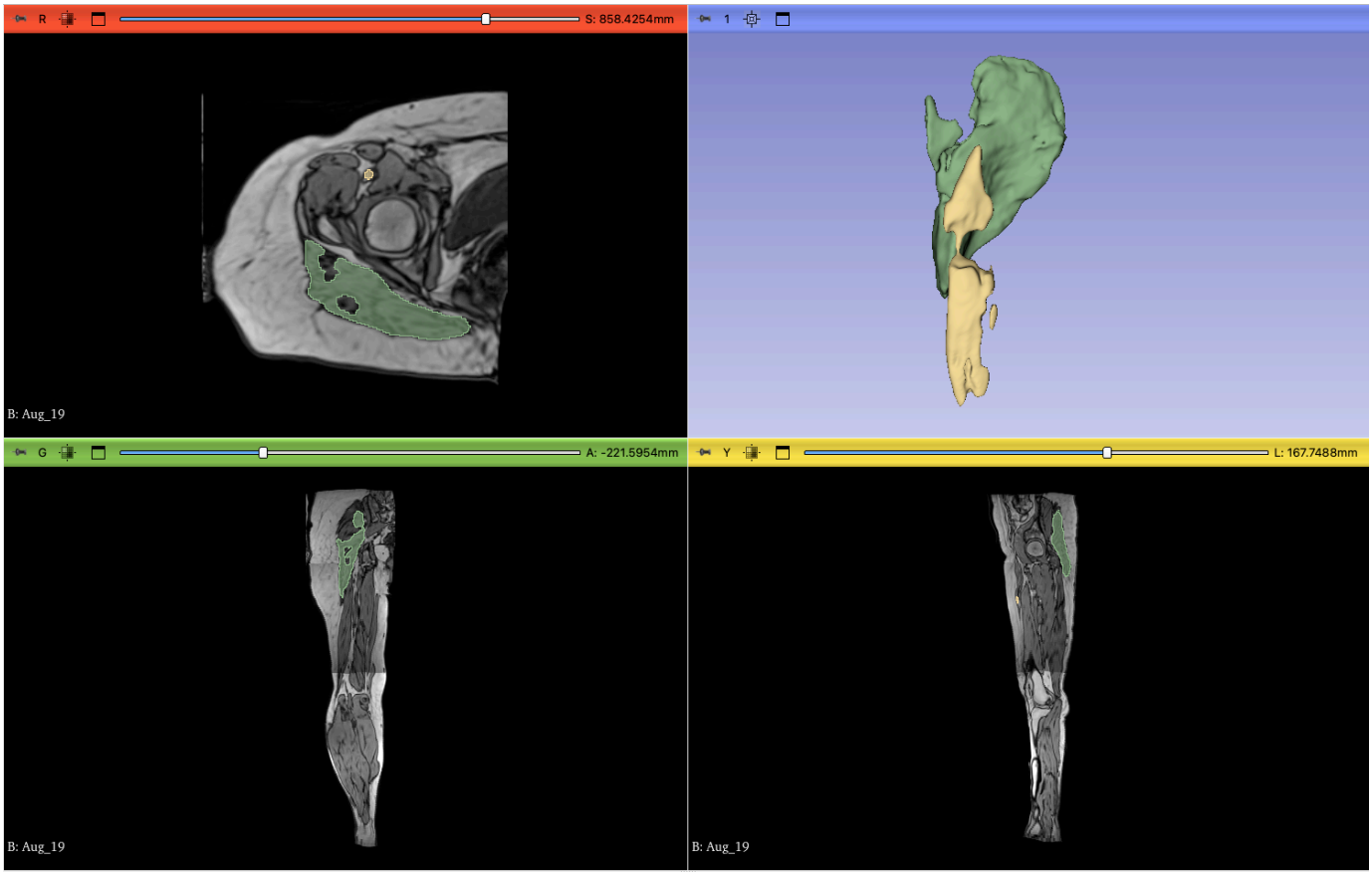


Fig. 19: Muscles in the MRI data from Henson et al. [24] were predicted using the BODIES_Muscles021 model. It showed decreased accuracy due to differences in MRI modality, but true performance, measured by DSC and HD, was not meaningfully evaluated, as the available 'ground truth' segmentations were of poor quality. Nevertheless, for broader applicability, both the model and dataset should likely be trained on multiple MRI modalities.

APPENDIX K
SEGMENTATION OF LOWER BODY PER SUBJECT

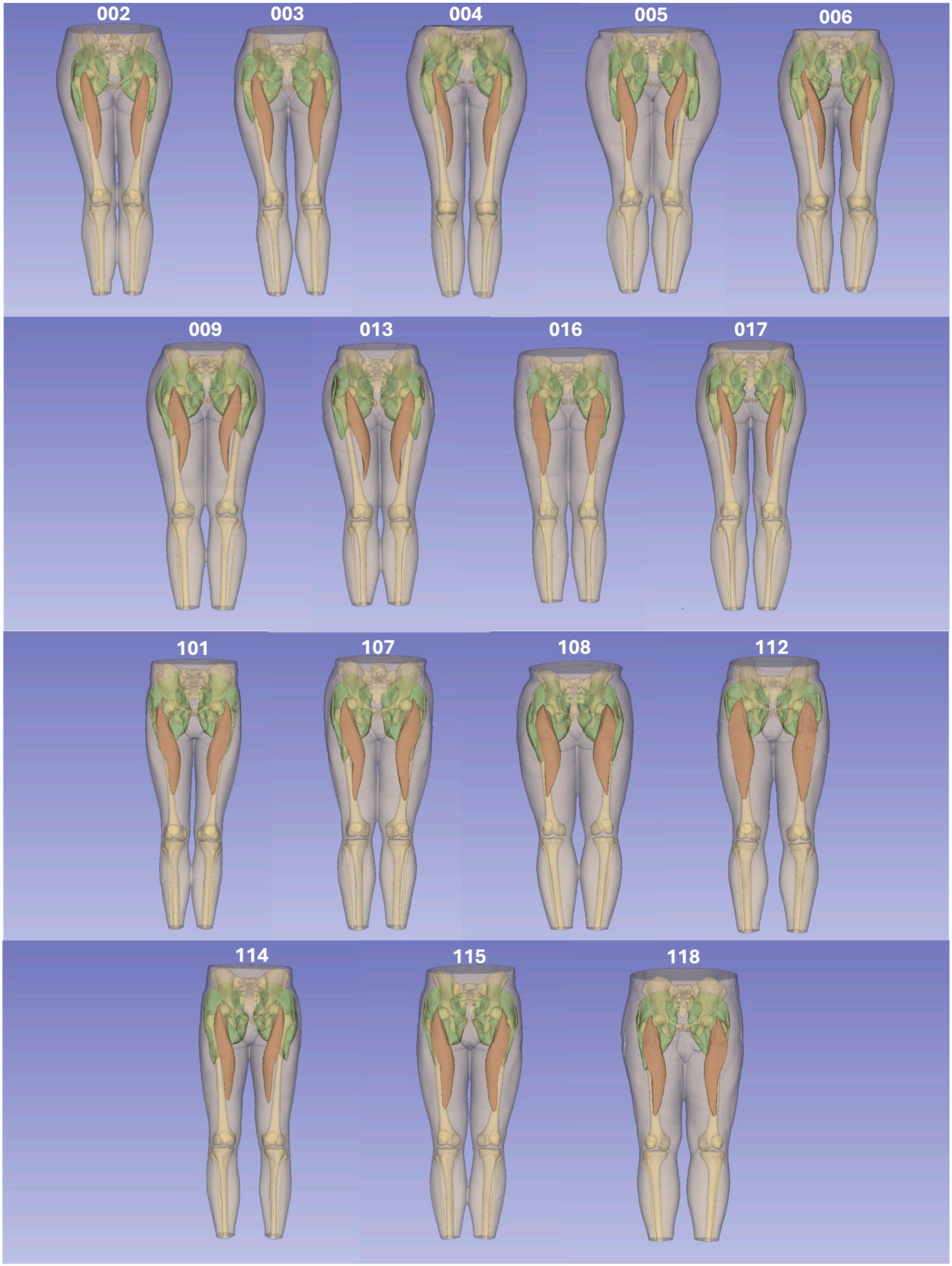


Fig. 20: Anterior view of the segmented lower extremity in all subjects, displaying body shape, skeleton, and muscles of interest (green: GMAX, red/brown: RFEM). Female subjects are illustrated on the upper half and male subjects on the lower half.

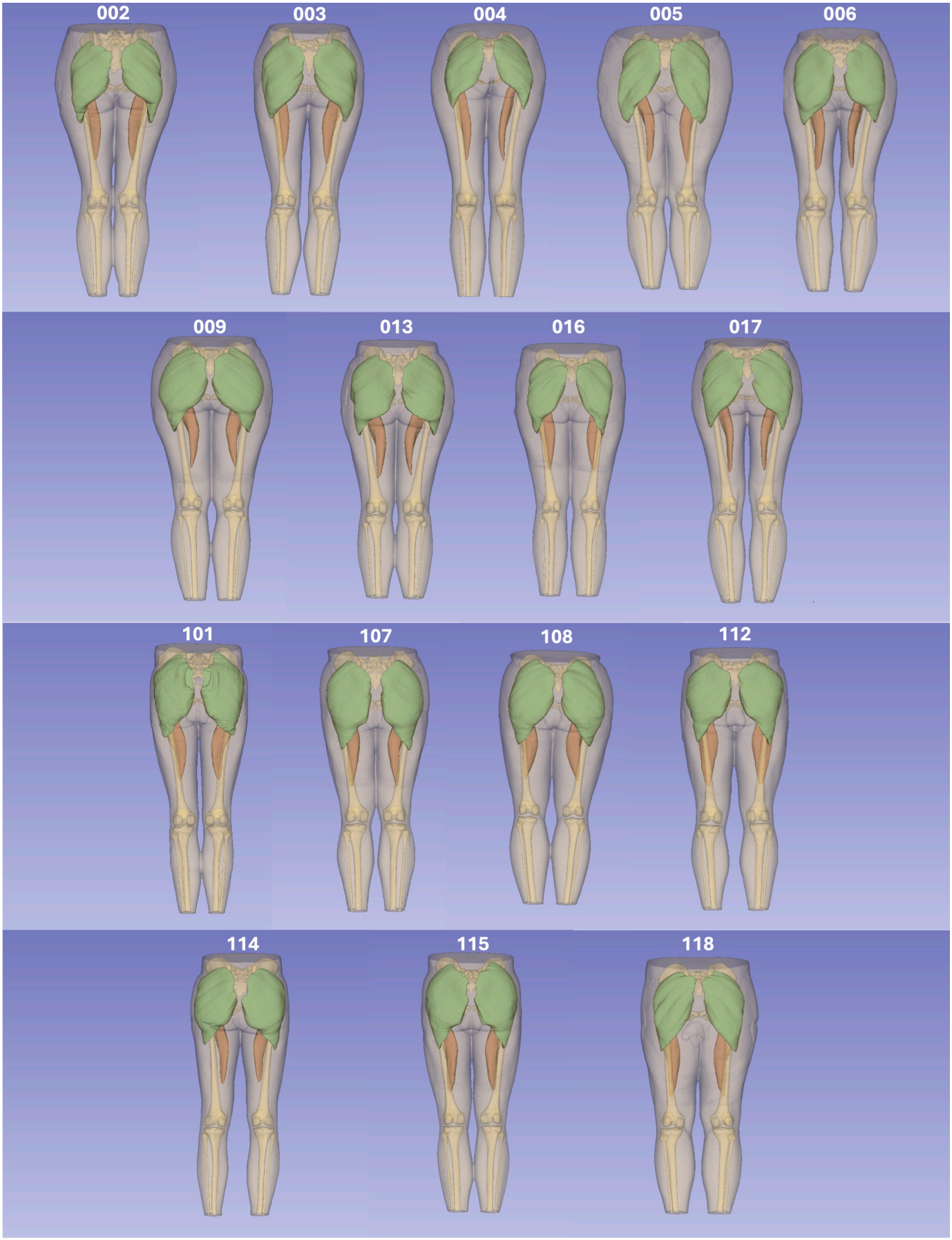


Fig. 21: Posterior view of the segmented lower extremity in all subjects, displaying body shape, skeleton, and muscles of interest (green: GMAX, red/brown: RFEM). Female subjects are illustrated on the upper half and male subjects on the lower half.

APPENDIX L
PELVIC GEOMETRY COMPARISON

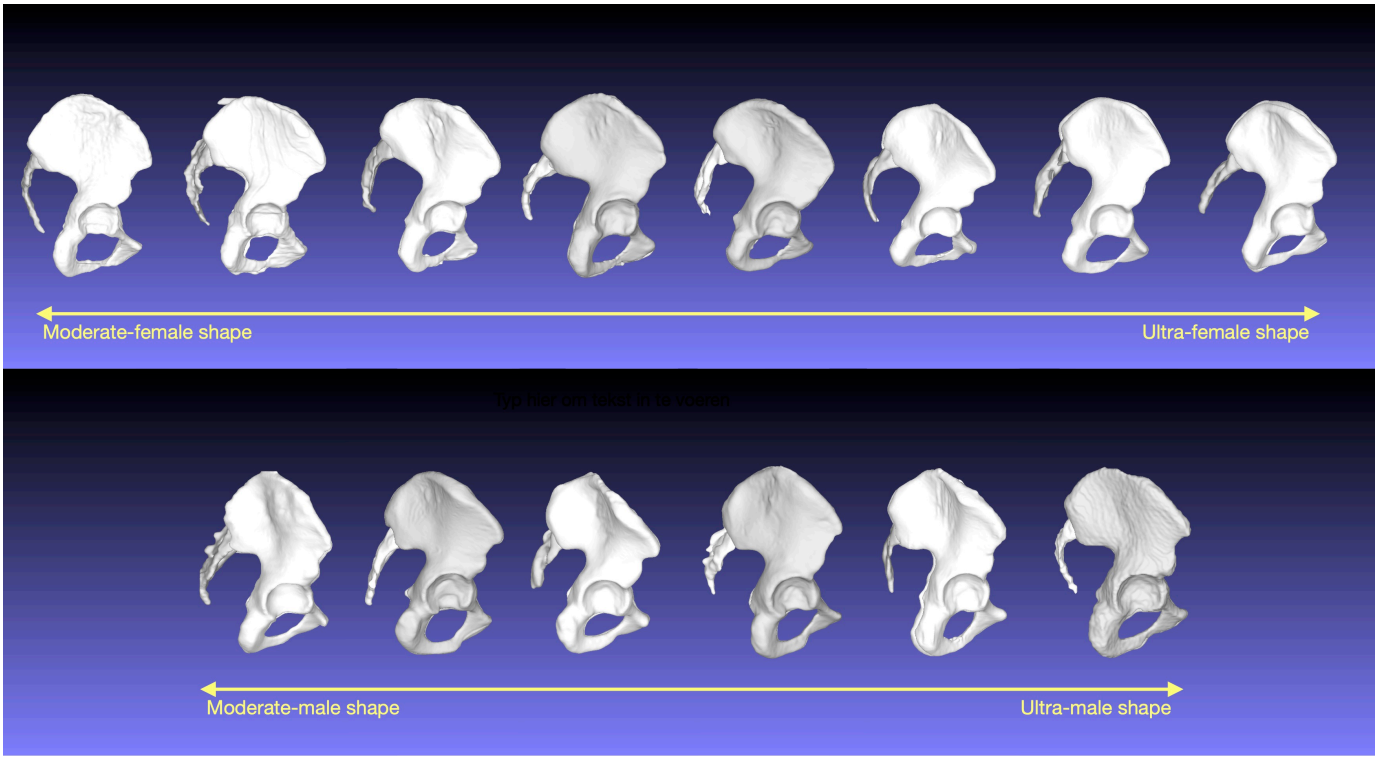


Fig. 22: All pelves in this study illustrated from the sagittal plane were ordered based on their shape. The "ultra-female shape" showed a more forward-tilted pelvis with an outward and posterior pointing sacrum. On the other hand, the "ultra-male shape" displayed a more straight-up ilium together with a curved and pointing inwards sacrum.

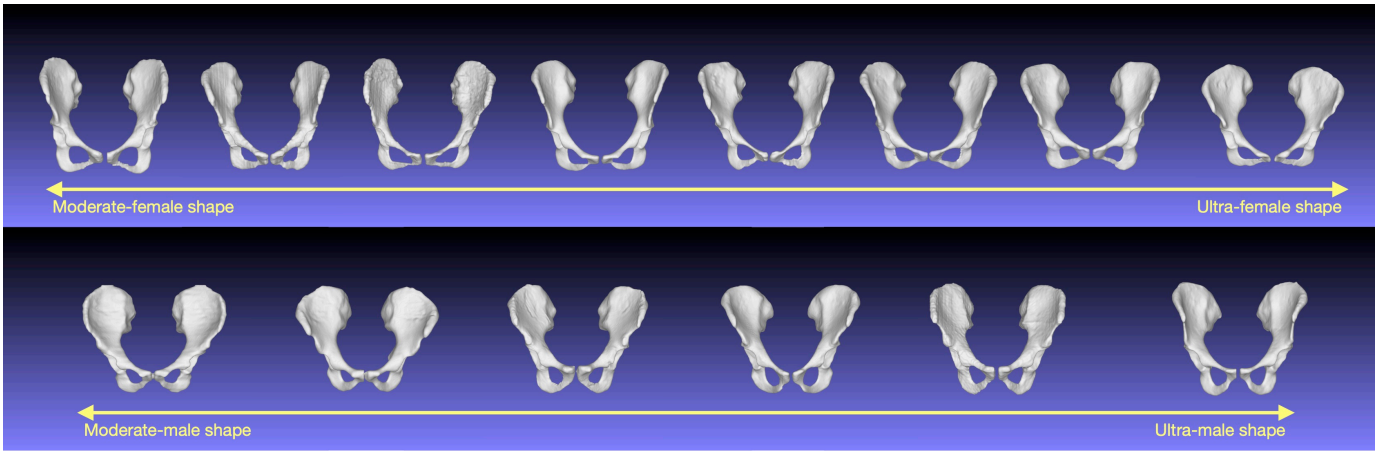


Fig. 23: All pelves in this study illustrated from the front view, order on there sexual dimorphism according to Fischer et al. [17].

APPENDIX M
MUSCLE MOMENT ARMS

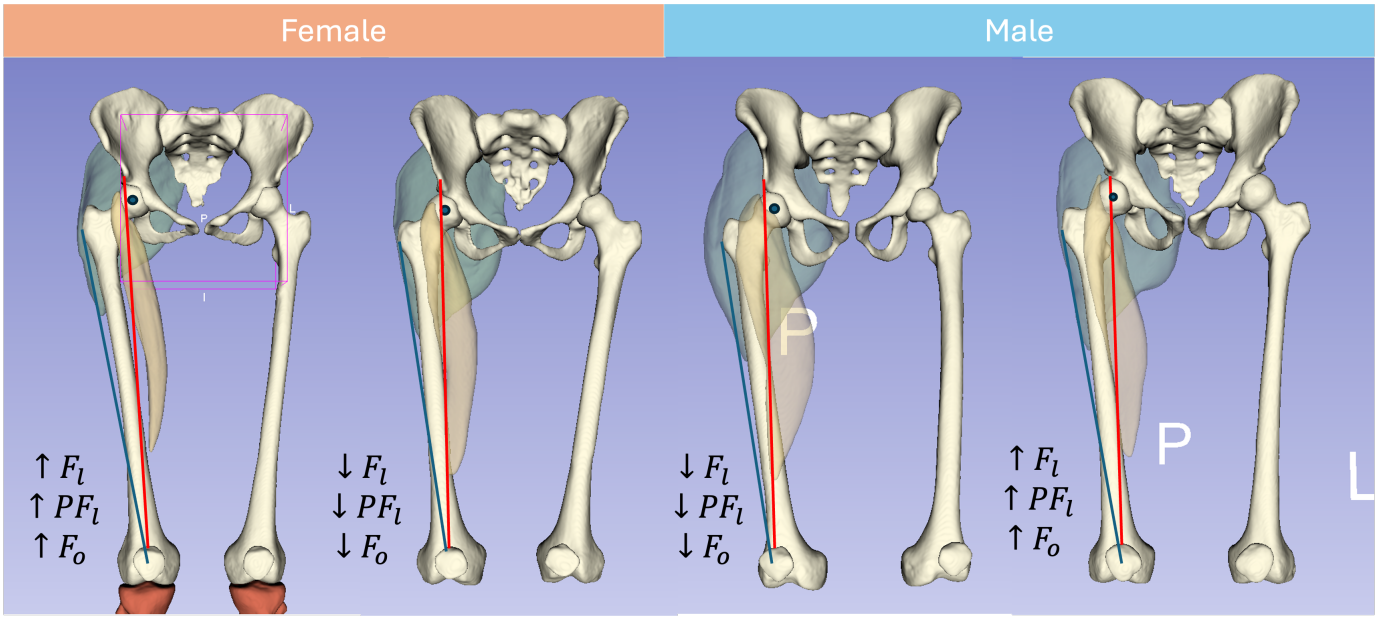


Fig. 24: Line-of-actions for RFEM from frontal view based on subjects from our dataset that should relative differences in the predictive bone variables. Red lines indicate the line-of-action for RFEM, and in yellow, their moment arms and the dot represent the hip joint center

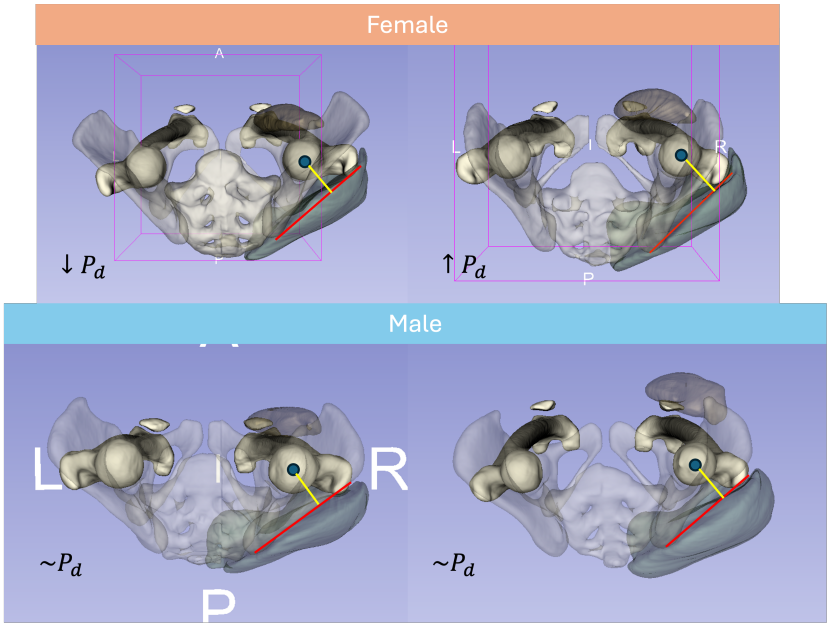


Fig. 25: Moment arms for hip exorotation for GMAX based on subject from our dataset that should relative differences in the predictive bone variables. Red lines indicate the line-of-action for GMAX, and in yellow, their moment arms and the dot represent the hip joint center.