

Article

AI-Enhanced Lower Extremity X-Ray Segmentation: A Promising Tool for Sarcopenia Diagnosis

Hyunwoo Park ¹, Hyeonsu Kim ² and Junil Yoo ^{3,*}¹ Department of Internal Medicine, Hallym University Medical Center, Hallym University College of Medicine, Anyang 14068, Republic of Korea; andyr2d2@hallym.or.kr² Department of Biomedical Research Institute, Inha University Hospital, Incheon 22188, Republic of Korea; lemonjames96@gmail.com³ Department of Orthopedic Surgery, Inha University Hospital, Inha University College of Medicine, Incheon 22188, Republic of Korea

* Correspondence: furim@daum.net

Abstract

Background/Objectives: Sarcopenia, characterized by progressive loss of skeletal muscle mass and strength, significantly impacts physical function and quality of life in older adults. Traditional measurement methods like Dual-energy X-ray absorptiometry (DEXA) are often inaccessible in primary care. This study aimed to develop and validate an AI-driven auto-segmentation model for muscle mass assessment using long X-rays as a more accessible alternative to DEXA. **Methods:** This was a retrospective validation study using data from the Real Hip Cohort at Inha University Hospital in South Korea. 351 lower extremity X-ray images from 157 patients were collected and analyzed. AI-based semantic segmentation models, including U-Net, V-Net, and U-Net++, were trained and validated on this dataset to automatically segment muscle regions. Model performance was assessed using Intersection over Union (IoU) and Dice Similarity Coefficient (DC) metrics. The correlation between AI-derived muscle measurements and the DEXA-derived skeletal muscle index was evaluated using Pearson correlation analysis and Bland–Altman analysis. **Results:** The study analyzed data from 157 patients (mean age 77.1 years). The U-Net++ architecture achieved the best segmentation performance with an IoU of 0.93 and DC of 0.95. Pearson correlation demonstrated a moderate to strong positive correlation between the AI model's muscle estimates and DEXA results ($r = 0.72$, *** $p < 0.0001$). Regression analysis showed a coefficient of 0.74, indicating good agreement with reference measurements. **Conclusions:** This study successfully developed and validated an AI-driven auto-segmentation model for estimating muscle mass from long X-rays. The model provides an accessible alternative to DEXA, with potential to improve sarcopenia diagnosis and management in community and primary care settings. Future work will refine the model and explore its application to additional muscle groups.

Keywords: deep learning; semantic segmentation; X-ray; thigh; skeletal muscle



Academic Editor: Gianluigi Pasta

Received: 5 August 2025

Revised: 22 September 2025

Accepted: 26 September 2025

Published: 30 September 2025

Citation: Park, H.; Kim, H.; Yoo, J. AI-Enhanced Lower Extremity X-Ray Segmentation: A Promising Tool for Sarcopenia Diagnosis. *Healthcare* **2025**, *13*, 2488. <https://doi.org/10.3390/healthcare13192488>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sarcopenia, characterized by progressive loss of skeletal muscle mass and strength, has emerged as one of the most critical chronic conditions in recent decades, with its significance extending far beyond mere physical decline to profoundly impact quality of life, morbidity of chronic diseases, and remaining life expectancy of elderly populations [1–3]. As global aging populations continue to expand, sarcopenia has evolved from a recognized

aging phenomenon into a major public health challenge, affecting 10–27% of individuals over 60 years and up to 30% of those over 65 years, though prevalence varies considerably across countries and diagnostic criteria [4]. For instance, South Korea reports prevalence rates of 6.6% in men and 10.3% in women aged 65 and older, while rural areas consistently show higher rates than urban regions due to differences in lifestyle, nutrition, and healthcare accessibility [5]. The socioeconomic burden is staggering, with direct healthcare costs reaching approximately \$18.5 billion annually in the United States and £2.5 billion in the United Kingdom, translating to additional costs of \$2315.7 per patient in the US and £2707 per patient in the UK [6]. This economic impact is particularly pronounced among socioeconomically disadvantaged populations, who experience higher incidence rates and reduced access to early prevention and treatment, thereby exacerbating health inequalities [7].

Over the past 30 years, continental consensus groups have developed and refined diagnostic criteria, culminating in the recent Global Leadership Initiative on Sarcopenia (GLIS) guidelines, which have narrowed the focus to precise muscle mass measurement and specific muscle strength assessments [8]. However, despite these advances, significant challenges persist in clinical practice, including lack of standardization across different diagnostic criteria, insufficient availability of specialized equipment and trained personnel for accurate muscle mass and strength evaluation, and limited awareness among healthcare providers regarding the importance of early sarcopenia detection and management [9]. While effective management requires multidisciplinary approaches integrating nutrition, exercise, and pharmacological interventions, the fundamental barrier lies in the inconsistent and often inadequate diagnostic practices that prevent timely identification and appropriate therapeutic intervention [10,11].

Previous guidelines, such as those from the Asian Working Group for Sarcopenia (AWGS 2019) and the European Working Group for Sarcopenia in Older People (EWG-SOP2), established standards for identifying and assessing sarcopenia based on three aspects: muscle mass (skeletal muscle index), muscle strength (hand grip strength), and physical performance such as gait speed and short physical performance battery (SPPB) [12,13]. However, in the process of developing the GLIS guidelines, sarcopenia has been defined in terms of muscle mass, muscle strength, and muscle-specific strength (muscle strength/muscle size), underscoring the critical importance of accurate skeletal muscle mass measurement. Currently, the gold standard methods for measuring muscle mass include whole-body Dual-energy X-ray absorptiometry (DEXA), CT and MRI [14–16]. However, these advanced imaging techniques present significant challenges in primary care settings due to their limited accessibility [17]. While Bioelectrical Impedance Analysis (BIA) could be an alternative, its lack of visual representation of muscle tissue limits its ability to evaluate the effectiveness of specific exercise interventions or treatments based solely on numerical changes [18,19]. Therefore, the need for more accessible, user-friendly diagnostic devices, such as X-ray and ultrasound, is being emphasized, especially considering that chronic conditions like sarcopenia are ideally managed at the community and primary care settings.

Recent advances in artificial intelligence have revolutionized medical imaging analysis for muscle mass assessment, particularly in CT and MRI interpretation. Studies have demonstrated that AI-based approaches offer unique advantages including elimination of inter-observer variability that commonly affects manual measurements, capability to detect subtle radiographic patterns invisible to human observers, standardization of measurement protocols across different healthcare settings, and potential for continuous learning through exposure to diverse patient populations [20]. Recent research published in Scientific Reports has demonstrated the effectiveness of U-net transformer architecture for pre-

cise individual muscle segmentation in whole thigh CT scans for sarcopenia assessment, achieving superior accuracy in muscle boundary delineation and quantification compared to conventional manual methods [21].

Given these promising developments in AI-assisted muscle assessment and the pressing need for more accessible diagnostic tools in primary care, X-ray imaging emerges as an ideal candidate for AI-driven muscle analysis due to its widespread accessibility in primary healthcare settings. Therefore, the purpose of this study is to develop and validate a novel AI-driven automated segmentation model specifically designed for muscle mass assessment using whole-body X-rays, with the aim of establishing a standardized imaging protocol and demonstrating clinical validity through direct comparison with gold standard DEXA measurements.

2. Methods

2.1. Introduction

This study aimed to develop and validate a robust AI-based semantic segmentation model for accurately quantifying muscle mass in the thigh and calf regions from lower extremity X-ray images. Leveraging a comprehensive dataset comprising 351 X-ray images and 66 paired Dual-Energy X-ray Absorptiometry (DEXA) scans from the Real Hip Cohort in South Korea, the methodology involved the development and training of advanced deep learning architectures, specifically U-Net, V-Net, and U-Net++, on expertly annotated and pre-processed images. The developed model's performance was rigorously evaluated using various metrics, including Intersection over Union (IoU), Dice Coefficient (DC), Average Distance (AD), and Hausdorff Distance (HD), with its accuracy and reliability subsequently validated through statistical comparisons, such as Pearson's correlation coefficient and Bland-Altman analysis, against DEXA-derived skeletal muscle index measurements. All research procedures strictly adhered to the principles of the Declaration of Helsinki and received approval from the Institutional Review Board of Inha University Hospital (IRB No. INUH 2023-04-027).

2.2. Study Design

This study developed a robust AI-based semantic segmentation model to accurately measure muscle groups in the thigh and calf regions. Furthermore, a comparative study against Dual-Energy X-ray Absorptiometry (DEXA) scans was conducted to validate the accuracy and reliability of the AI-based semantic segmentation model for quantifying muscle mass. The training process for this model utilized a dataset comprising 351 lower extremity X-ray images from the Real Hip Cohort, a comprehensive hip fracture cohort based in South Korea.

This study adhered to the principles of the Declaration of Helsinki and was approved by the Institutional Review Board (IRB No. INUH 2023-04-027) at Inha University Hospital. All research procedures were carried out with strict adherence to ethical standards, including protection of participant privacy, confidentiality, and rights.

2.3. Study Setting and Participants

The study assembled a cohort of 157 patients from the Real Hip Cohort at Inha University Hospital in South Korea. This cohort provides a comprehensive repository of medical imaging data and clinical information. Specifically, 351 lower extremity X-ray images and 66 Dual-Energy X-ray Absorptiometry (DEXA) scans were collected from these individuals.

Initially, 878 individuals were identified, from whom 157 had lower extremity X-ray data (351 images). Among these, 351 images were ultimately used for semantic segmenta-

tion (281 for training and 70 for validation), after excluding 72 single-leg rotation images and 60 external rotation images due to rotation. For outcome analysis, 62 individuals provided 66 paired DEXA scans and X-ray images, and 1-year follow-up data from 4 individuals (8 paired DEXA scans and X-ray images) were also included for further discussion. The specific criteria for this study and the patient selection flowchart are detailed in Figure 1.

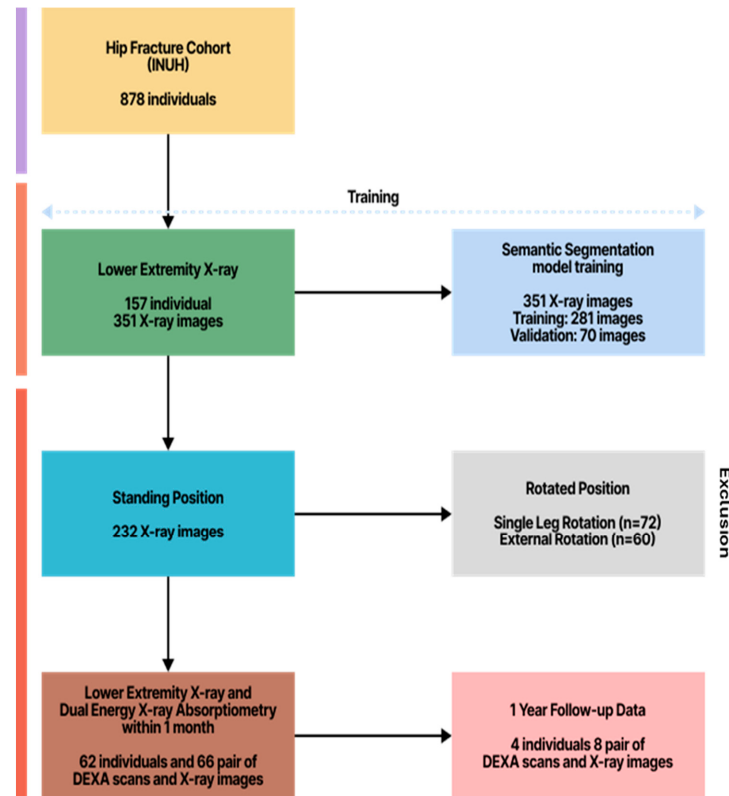


Figure 1. Patient selection flowchart. From 878 initial participants, 157 provided X-ray data (351 images). After excluding rotated images, 351 images were used for AI model development (281 training, 70 validation). Sixty-six paired DEXA-X-ray datasets from 62 participants were used for validation analysis.

2.4. Data Acquisition and Ground Truth Labeling

Each participant's lower extremity X-ray images were taken in a standing position using a X-ray machine (Siemens Healthineers, Erlangen, Germany). These images capture the entire lower extremities, spanning from the pelvic region down to the ankles, and are stored in the industry-standard DICOM file format. Participants whose lower extremities were externally rotated were excluded from the training process, as this rotation could affect the accuracy of the AI-based segmentation model by causing changes in distinct regions (lateral, medial, and calf) of the X-ray, particularly affecting the lower leg.

Additionally, Dual-Energy X-ray Absorptiometry (DEXA) scans were performed under a standard protocol using a GE Lunar DXA machine (GE Healthcare, Madison, WI, USA). These scans provided a detailed assessment of bone density and body composition, including detailed measurements of fat mass (g) and lean mass (g) for various body regions like arms, legs, and trunks.

To establish ground truth labels for training the AI model, we employed an annotation tool provided by 3D Slicer software (version 5.6.2). This allowed for precise delineation of various anatomical structures and tissue types. As displayed in Figure 2, we segmented the muscles into three distinct classes: lateral, medial, and gluteal regions within the thigh, as well as the calf region. Furthermore, we annotated subcutaneous fat and bony structures

such as the femur, tibia, and iliac bones. These meticulously labeled annotation masks were saved in the NRRD file format, which incorporates crucial information regarding pixel spacing in real-world units, enabling accurate quantification and analysis.

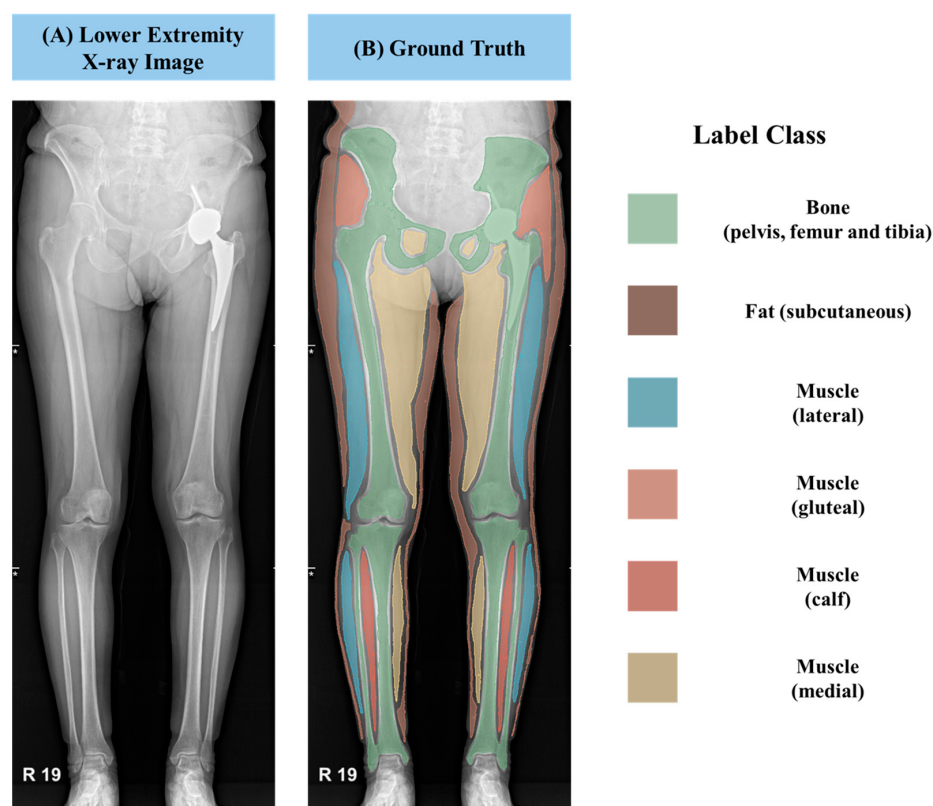


Figure 2. (A) Lower extremity X-ray image and (B) Corresponding ground-truth segmentation mask with color-coded labels for bone (pelvis, femur, and tibia), subcutaneous fat, and multiple muscle regions (lateral, posterior, gluteal, calf, medial). These labeled classes form the basis for subsequent segmentation and analysis.

For the manual segmentation, a standardized protocol was meticulously followed, utilizing anatomical landmarks and boundary definitions established based on the Netter Atlas of Human Anatomy. This rigorous approach ensured consistency and accuracy across all images. The segmentation process incorporated a rigorous three-tier quality control procedure: initially, the first author (H. Park, Department of Internal Medicine) performed the primary manual segmentation of all images. Subsequently, a co-author (H. Kim, Department of Biomedical Research Institute) independently double-checked all segmented images for accuracy and strict adherence to the anatomical guidelines. This co-author brings extensive experience in developing AI models for muscle mass measurement using lower extremity CT scans [21]. Finally, the corresponding author (J. Yoo, Department of Orthopedic Surgery), a board-certified orthopedic specialist with profound expertise in musculoskeletal anatomy and imaging, provided final confirmation for all segmented images. This structured approach, leveraging multiple expert reviews, was instrumental in ensuring the high quality, consistency, and clinical relevance of the ground truth segmentation datasets for AI model training.

2.5. Deep-Learning Methods of Semantic Segmentation Model

2.5.1. Pre-Processing and Augmentation

In order to augment the relatively limited training dataset, we randomly flipped the image horizontally. This augmentation technique helps increase the diversity of the train-

ing data and improve the model's ability to generalize. The images were resized to (512, 1024) Numpy Shape for training the model, ensuring consistent input dimensions for the U-Net based neural network.

Additionally, we applied several pre-processing steps using image transforms to segments muscle tissue, subcutaneous fat and skeletal bone:

1. **Scale Intensity** (min = 0.0, max = 1.0): This transform scales the intensity values of the image to a range between 0.0 and 1.0. Normalizing the intensity range helps standardize the input data across different images, which can improve the model's learning process and performance.
2. **Histogram Normalize**: Histogram normalization is applied to enhance image contrast. This technique redistributes the intensity values of the image to convert the full range of possible intensities more evenly. It can help in making features more distinguishable, especially in medical imaging where contrast can be critical for identifying structures.
3. **Normalize Intensity**: Standardized input data by subtracting the mean and dividing by the standard deviation of image intensities, aiding faster convergence and improved performance.
4. **Adjust Contrast** (gamma = 1.6): Gamma correction was used to increase contrast, making darker regions darker and brighter regions brighter, which enhances image features.

As shown in Figure 3, these pre-processing steps work together to standardize the input data, enhance important features and improve the overall quality of the images before they are fed into the neural network for training. Crucially, all these pre-processing procedures, including image resizing and random horizontal flipping, were rigorously standardized. Each step was applied consistently to every image using the same predefined algorithms and parameters (e.g., Scale Intensity to min = 0.0, max = 1.0; Adjust Contrast with gamma = 1.6). This entire process was fully automated and executed via programmatic scripts, ensuring an entirely objective approach without any human intervention or subjective judgment.

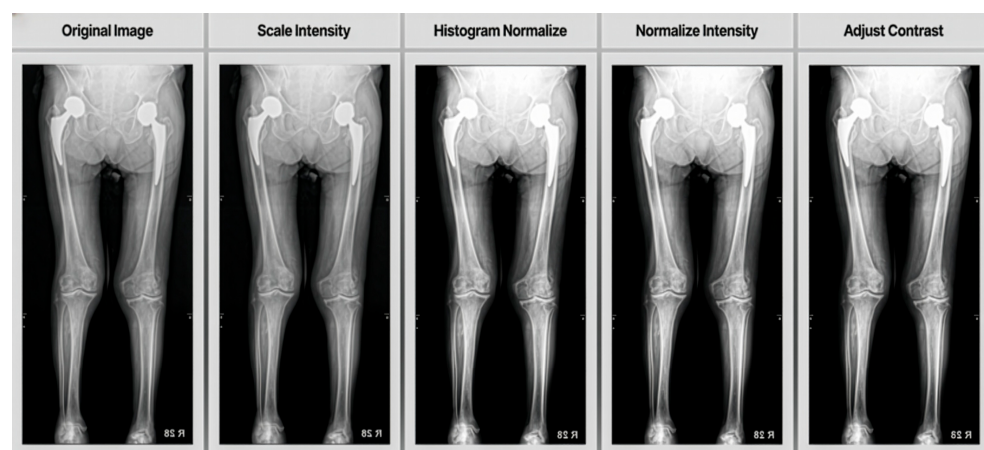


Figure 3. To prepare the X-ray images for segmentation, four key pre-processing steps were applied: intensity values were first scaled to a 0.0–1.0 range to ensure consistent brightness; histogram normalization redistributed pixel intensities for enhanced contrast; subsequent intensity normalization (subtracting the mean and dividing by the standard deviation) further standardized the data; and finally, gamma correction ($\gamma = 1.6$) adjusted contrast to emphasize important anatomical features.

2.5.2. Model Architectures

For the AI-based automatic segmentation model, we employed U-Net, V-Net, and U-Net++ architectures. U-Net is known for its U-shaped encoder–decoder structure with skip connections, making it effective for biomedical image segmentation by combining low-level and high-level features for precise object boundary localization [22]. V-Net extends U-Net for 3D image segmentation, utilizing 3D convolutional layers and incorporating residual learning to facilitate the training of deeper networks, with strided convolutions for downsampling to preserve spatial information [23]. U-Net++, an advanced version of U-Net, further improves segmentation accuracy through nested and dense skip connections that reduce the semantic gap between encoder and decoder feature maps, along with deep supervision to enhance gradient flow [24]. These architectures were chosen for their proven capabilities in various medical imaging segmentation tasks.

2.5.3. Model Training

Training was conducted on high-performance workstations equipped with NVIDIA RTX 4090 and RTX A6000 GPUs, operating on an Ubuntu 20.04 platform. The deep learning-based segmentation model was implemented using PyTorch version 2.4.1 (<https://pytorch.org/>) and MONAI version 1.3.2 (<https://monai.io/>) frameworks.

The Dice coefficient was adopted as the loss function, a frequent choice in semantic segmentation. The model was fine-tuned using the AdamW optimizer with the following hyperparameters: learning rate: 8×10^{-5} , weight decay: 1×10^{-5} , batch size: 8, and 3000 epochs. These hyperparameters were determined through a grid search methodology, exploring learning rates from 1×10^{-3} to 1×10^{-5} , with the final configuration selected via heuristic techniques due to its superior performance on the muscle segmentation datasets.

2.6. Performance Evaluation and Statistical Analysis

To evaluate our model's performance comprehensively, we employed a range of metrics: (i) Intersection over Union (IoU), (ii) Dice Coefficient (DC), (iii) Average Distance (AD), (iv) Hausdorff Distance (HD), and (v) Relative Absolute Area Difference (RAAD).

The IoU and DC provide clear indicators of overlap accuracy between the predicted and ground truth segments, essential for assessing overall segmentation quality. The AD measures the mean spatial discrepancy between these boundaries, highlighting the model's precision in contour delineation. The HD focuses on the maximum boundary discrepancy, emphasizing the model's capability to capture the most significant boundary deviations, which is critical for precise applications. Finally, the RAAD evaluates the proportional difference in area, offering insight into the model's consistency in estimating the size of segmented regions. Each metric contributes unique insights and addresses specific aspects of the segmentation process, facilitating a thorough and nuanced assessment of model performance.

(i) Intersection over Union (IoU) for Multiclass

For each class c within the total number of classes n , the IoU was calculated by identifying the number of pixels where both the prediction and the ground truth label were equal to c (intersection), and the number of pixels where either the prediction or the ground truth label was c (union). The IoU score for each class c was computed using the formula

$$\text{IoU}_c = \frac{\sum \text{pred} \wedge \text{label}}{\sum \text{pred} \vee \text{label}} \quad (1)$$

where pred and label are binary masks indicating the presence of class c in the predicted and ground truth label images, respectively. If the union of the predicted and ground

truth pixels for a class was zero, the IoU was set to 1, indicating a perfect match. The overall performance was quantified by averaging the IoU scores across all classes:

$$\text{Mean IoU} = \frac{1}{n} \sum_{c=0}^{n-1} \text{IoU}_c \quad (2)$$

(ii) Dice Coefficient (DC) for Multiclass

The Dice Coefficient, another metric to assess segmentation accuracy, was calculated for each class c by determining the number of pixels where both the predicted and ground truth labels were c (intersection) and the total number of pixels predicted as c or labeled as c (union). The Dice score for each class was calculated as follows:

$$\text{Dice}_c = \frac{2 \times \Sigma(\text{pred} \cdot \text{label})}{\Sigma(\text{pred}) + \Sigma(\text{label})} \quad (3)$$

The final Dice Coefficient across all classes was computed by averaging the individual class scores:

$$\text{Mean Dice} = \frac{1}{n} \sum_{c=0}^{n-1} \text{Dice}_c \quad (4)$$

(iii) Average Distance (AD) for Multiclass

The Average Distance (AD) metric was utilized to quantify the spatial discrepancy between predicted and ground truth segmentations. For each class c , the AD was determined by calculating the mean distance of each pixel in the predicted mask to the nearest pixel in the ground truth mask and vice versa. This was performed using the Euclidean distance transform on the complements of the binary masks:

$$\text{AvgDist}_c = \text{mean}(\text{pred}[\text{label}]) + \frac{\text{mean}(\text{label}[\text{pred}])}{2} \quad (5)$$

If both pred class and label class were empty, the average distance was set to 0. In cases where one was empty and the other was not, the distance was considered infinite. The overall AD was averaged across all classes:

$$\text{Mean AvgDist} = \frac{1}{n} \sum_{c=0}^{n-1} \text{AvgDist}_c \quad (6)$$

(iv) Relative Absolute Area Difference (RAAD) for Multiclass

The Relative Absolute Area Difference (RAAD) metric measured the percentage difference in area between the predicted and ground truth segments for each class c . This was calculated by determining the absolute difference between the areas of predicted and labeled pixels, normalized by the labeled area:

$$\text{RelAbsDiff}_c = \left| \frac{\text{pred}_{\text{area}} - \text{label}_{\text{area}}}{\text{label}_{\text{abd}}} \right| \times 100\% \quad (7)$$

where $\text{pred}_{\text{area}}$ and $\text{label}_{\text{area}}$ denote the number of pixels predicted and labeled as class c , respectively. In instances where the $\text{label}_{\text{area}}$ was zero, the RAAD was set to 0 if $\text{pred}_{\text{area}}$ was also zero (indicating no discrepancy), or 100% otherwise. The final score was computed as a weighted average based on class areas:

$$\text{Weighted Avg RelAbsDiff} = \frac{\sum_{c=0}^{n-1} \text{RelAbsDiff}_c \times \text{class}_{\text{area}_c}}{\text{total}_{\text{area}}} \quad (8)$$

(v) Hausdorff Distance (HD) for Multiclass

The Hausdorff Distance (HD) metric was employed to measure the maximum boundary distance between predicted and ground truth segmentations for each class c . The HD for each class was computed by finding the maximum distance from any point in the predicted segmentation to the nearest point in the ground truth segmentation and vice versa:

$$\text{HausDorff}_c = \max(\text{directed_hausdorff}(\text{pred_points}, \text{label_points}), \text{directed_hausdorff}(\text{label_points}, \text{pred_points})) \quad (9)$$

where pred_points and label_points represent the coordinates of pixels in the predicted and ground truth masks for class c , respectively. If both pred and label were empty, the HD was considered 0. If one was empty and the other was not, the HD was set to infinity. The final Hausdorff distance was averaged across all classes:

$$\text{Mean Hausdorff} = \frac{1}{n} \sum_{c=0}^{n-1} \text{Hausdorff}_c \quad (10)$$

2.7. Comparison with Dual-Energy X-Ray Absorptiometry

2.7.1. DEXA Measurements

Dual-Energy X-ray Absorptiometry (DEXA) scans were performed on a subset of 66 subjects using Lunar Prodigy Advance DEXA machine. The skeletal muscle index was calculated from these scans according to standard protocols, defined as the lean mass (arms and legs) divided by the square of the subject's height (kg/m^2). DEXA measurements were conducted within [time frame] of the X-ray imaging to ensure comparability.

2.7.2. Comparison of Segmentation Results with DEXA

The muscle area derived from the segmentation models was compared with the DEXA skeletal muscle index. Muscle area from segmentation was calculated by summing the pixels classified as muscle and converting them to physical area using known pixel dimensions. Statistical analysis was performed using Python (version 3.9) with the SciPy and statsmodels libraries.

Pearson correlation analysis was conducted to assess the linear relationship between the segmentation-derived muscle area and the DEXA skeletal muscle index. Bland–Altman plots were generated to visualize the agreement between the two measurement methods, showing the mean difference and 95% limits of agreement. Paired t-tests were performed to determine if there were significant differences between the segmentation-based estimates and DEXA measurements.

3. Results

3.1. Demographic Characteristics

The study included 157 patients, with a demographic breakdown of 117 females and 40 males, and an average age of 77.1 years (77.16 years for females and 77.04 years for males). To evaluate the correlation between segmented muscle regions and skeletal muscle in the lower limbs, we utilized a total of 66 lower extremity X-ray images and 66 DEXA scans from 62 patients. Among the 66 DEXA results, 19 patients were classified as non-sarcopenic, 8 as having sarcopenia, and 39 as having severe sarcopenia.

A detailed overview of the participants' demographic data, including the number of participants, age, height, and weight (mean $\pm$ SD) for female, male, and total participants across non-sarcopenia, sarcopenia, and severe sarcopenia groups (as determined by DEXA measurements), is presented in Table 1. In addition to imaging data, the Real Hip Cohort provides extensive medical data, including physical performance met-

rics such as the Short Physical Performance Battery (SPPB), Koval's score, grip strength, and gait analysis data, which allowed for a comprehensive evaluation of the participants' musculoskeletal condition.

Table 1. Characteristics of the participants in DEXA measured groups.

		Female	Male	Total
Non-Sarcopenia	Number	19	-	19
	Age (SD)	79.9 (8.4)	-	79.9 (8.4)
	Height (SD)	156.0 (10.2)	-	156.0 (10.2)
	Weight (SD)	53.6 (9.6)	-	53.6 (9.6)
Sarcopenia	Number	7	1	8
	Age (SD)	79.2 (8.6)	83 (0)	79.6 (8.9)
	Height (SD)	153.3 (8.1)	169 (0)	156.3 (9.9)
	Weight (SD)	53.9 (7.7)	44 (0)	54.5 (9.4)
Severe Sarcopenia	Number	27	12	39
	Age (SD)	80.3 (8.3)	79.2 (7.7)	80.3 (8.3)
	Height (SD)	156.3 (10.1)	156.2 (10.5)	156.3 (10.0)
	Weight (SD)	54.6 (9.3)	54.7 (9.5)	53.6 (9.4)

Demographic data (number of participants, age, height, and weight [mean $\pm$ SD]) for female, male, and total participants are presented across non-sarcopenia, sarcopenia, and severe sarcopenia groups, as determined by DEXA measurements.

3.2. Semantic Segmentation Model Performance Evaluation

For the semantic segmentation model training, we employed 351 lower extremity X-ray images. As displayed in Table 2, our proposed model, which is based on the U-Net++ architecture, demonstrated superior performance in comparison to other models such as U-Net and V-Net. Specifically, the U-Net++ model achieved an Intersection over Union (IoU) of 0.93 and a Dice coefficient (DC) of 0.95, indicating a high degree of overlap between the predicted and ground truth segmentations.

Table 2. Performance of Semantic Segmentation Models in Validation Datasets.

	IoU	DC	AD	HD	RAAD (%)
U-Net	0.84	0.87	3.14	7.21	11.2
V-Net	0.87	0.91	2.87	6.50	7.13
U-Net++	0.93	0.95	0.89	1.92	1.80

U-Net, V-Net, and U-Net++ were evaluated using IoU, Dice Coefficient, Average Distance, Hausdorff Distance, and RAAD. U-Net++ achieved the highest IoU (0.93) and Dice (0.95) alongside the lowest Average Distance (0.89), Hausdorff Distance (1.92), and RAAD (1.80%).

Furthermore, U-Net++ also recorded the lowest Average Distance (AD) at 0.89 ± 0.021 , a Hausdorff Distance (HD) of 1.92 ± 0.101 , and a Relative Absolute Area Difference (RAAD) of $1.80 \pm 0.019\%$ among the evaluated models. These metrics collectively highlight U-Net++'s precise boundary delineation and minimal discrepancy in area measurements, thus underscoring its innovative and robust performance for semantic segmentation in lower extremity X-ray images. For comparison, U-Net achieved an IoU of 0.84 and a DC of 0.87, while V-Net showed an IoU of 0.87 and a DC of 0.91. Example images of the segmentation model output are displayed in the Supplementary Materials (Figure S1).

3.3. Validation of AI-Based Segmentation Model for Muscle Mass Quantification

To validate the accuracy of the AI-based semantic segmentation model for muscle mass quantification, we compared the results from the segmentation model with those obtained from the Dual-Energy X-ray Absorptiometry (DEXA) technique. The segmentation model's output was assessed against the DEXA skeletal muscle index, calculated as the appendicular lean mass divided by the square of the subject's height.

As presented in Table 3, the Pearson correlation coefficient between the total muscle region estimated by the segmentation model and the DEXA-derived skeletal muscle index was 0.72 ($*** p < 0.0001$), suggesting a moderate to strong positive correlation. Additionally, the correlation between the total muscle region estimated by the model and DEXA-measured legs' lean muscle mass was 0.66 ($*** p < 0.0001$). Individual muscle groups were further analyzed to evaluate the model's performance across different regions. Table 3 details the Pearson correlation coefficients for the medial, lateral, gluteal, and calf muscle groups against both the skeletal muscle index and legs' lean mass, indicating strong associations for most regions (e.g., medial 0.70 for skeletal muscle index and 0.57 for legs' lean mass; lateral 0.66 for skeletal muscle index and 0.72 for legs' lean mass; gluteal 0.65 for skeletal muscle index and 0.69 for legs' lean mass, all with $*** p < 0.0001$). The calf muscle showed a weaker correlation (0.10 for skeletal muscle index, 0.14 for legs' lean mass). Higher 'r' values indicate stronger associations between the X-ray-based segmentation results and DEXA-based indices.

Table 3. Pearson's correlation analysis between legs lean mass and skeletal muscle index from DEXA and muscle region from lower extremity X-ray.

	Skeletal Muscle Index (kg/m ²)		Legs Lean Mass (g)	
	Correlation Coefficient	p-Value	Correlation Coefficient	p-Value
Medial	0.70	*** <0.0001	0.57	*** <0.0001
Lateral	0.66	*** <0.0001	0.72	*** <0.0001
Gluteal	0.65	*** <0.0001	0.69	*** <0.0001
Calf	0.10	0.52	0.14	0.36
Total	0.72	*** <0.0001	0.66	*** <0.0001

***: $p < 0.001$. Pearson's correlation coefficients (r) and p-values comparing each muscle region identified by X-ray segmentation (medial, lateral, gluteal, calf, and total) with skeletal muscle index and legs lean mass derived from DEXA. Higher r values indicate stronger associations between the X-ray-based segmentation results and DEXA-based indices.

As depicted in Figure 4A, the Bland–Altman analysis revealed a mean difference of −2653 g, with 95% limits of agreement ranging from −8158 g to 2851 g, between the leg lean mass measured by DEXA and the estimated muscle region mass derived from lower extremity X-rays. This plot visually illustrates the agreement, or lack thereof, between the two measurement methods. Furthermore, an ordinary least squares regression analysis between the estimated muscle region and legs' lean mass produced a coefficient of 0.74 ($t = 28.87$, $p < 0.0001$), indicating a strong linear relationship between these two measurements, as shown in Figure 4B.

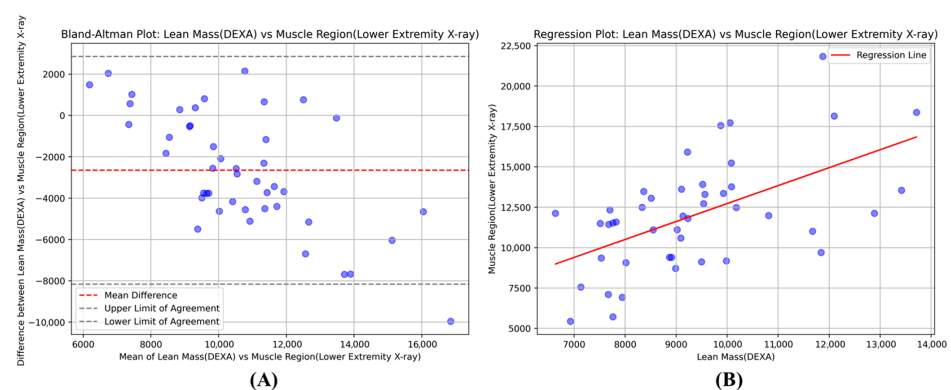


Figure 4. (A) The Bland–Altman plot illustrates the agreement between lean mass measured by DEXA and the muscle region identified on lower extremity X-rays, showing the mean difference (red dashed line) and upper/lower limits of agreement (black dashed lines). (B) The regression plot demonstrates the linear relationship between these two measurements, with the fitted regression line shown in red.

Finally, a paired t-test was conducted to assess the consistency between the muscle region estimated by the model and both the legs' lean mass and skeletal muscle index. The results indicated a ** significant difference between the model's estimated muscle region and legs' lean mass (t-statistic = -6.34 , $p < 0.05$), as well as between the estimated muscle region and skeletal muscle index ** (t-statistic = -23.77 , $p < 0.05$).

4. Discussion

4.1. Study Overview

This study developed and validated an AI-based semantic segmentation model to quantify muscle mass from lower extremity X-ray images as a potential alternative to DEXA for sarcopenia assessment. We compared U-Net++, U-Net, and V-Net architectures and evaluated the correlation between X-ray-derived muscle measurements and DEXA-based indices in 66 patients. The model achieved strong performance metrics with an IoU of 0.93 ± 0.015 and Dice coefficient of 0.95 ± 0.008 , demonstrating high accuracy in muscle boundary delineation. The correlation between X-ray-derived measurements and DEXA skeletal muscle index was $r = 0.72$ ($p < 0.0001$), indicating a statistically significant relationship between the two measurement methods.

4.2. Performance Evaluation and Clinical Significance

When compared to existing literature on muscle segmentation in lower extremity X-rays, our model's performance is competitive yet distinct in its approach. A comparable study reported a numerically higher IoU value of 0.959 (95% CI 0.959–0.960) than our achieved IoU of 0.93 [25]. However, that model demonstrated reduced accuracy in critical anatomical regions, particularly the medial and gluteal fold areas where precise segmentation is challenging. In contrast, our model was specifically designed to analyze five distinct muscle regions including the gluteal region, achieving moderate to strong correlations with DEXA measurements for four of the five regions. This comprehensive regional analysis represents a significant advancement over previous approaches, as it provides detailed information about individual muscle groups rather than treating the lower extremity as a single unit. While the calf region showed weaker correlation in our analysis, the overall multi-regional approach offers superior clinical utility for targeted muscle assessment and sarcopenia evaluation compared to existing single-measurement methodologies.

The weaker correlation observed specifically in the calf muscle group ($r = 0.10$) is hypothesized to be attributable to the inherent limitations of 2D X-ray projection imaging. Anatomically, the major posterior calf muscles (such as the Soleus and Gastrocnemius) are significantly overlapped and obscured by the denser anterior bone structures, particularly the tibia, when viewed on a single projection. This substantial bone-muscle overlap complicates the precise delineation of muscle boundaries, potentially leading to misclassification of muscle tissue during both the ground truth labeling and the subsequent AI segmentation. Nonetheless, this localized technical constraint does not diminish the overall efficacy of the model, given the strong correlations achieved in the functionally critical muscle groups—the medial thigh ($r = 0.70$), lateral thigh ($r = 0.66$), and gluteal region ($r = 0.65$). The model's robust performance in these three major regions underscores its potential as a reliable tool for sarcopenia assessment.

The correlation between X-ray-derived measurements and the DEXA skeletal muscle index was **statistically significant and moderate-to-strong ($r = 0.72$, $* p < 0.0001$). This relationship demonstrates that our AI-based method explains 52% of the variance ($r^2 = 0.52$) in muscle mass, capturing a substantial, clinically meaningful portion of the variation detected by the gold standard technique. However, the remaining unexplained variance of 48% necessitates careful clinical interpretation, especially as it suggests additional factors

may influence the relationship between the two measurements. This degree of correlation is particularly relevant for patients near sarcopenia diagnostic thresholds, where high measurement precision is critical for accurate clinical decision-making.

The Bland–Altman analysis provides essential insights into the clinical reliability and measurement agreement between our X-ray-based method and DEXA, particularly relevant for diagnostic applications. The systematic bias of -2653 g indicates that our model consistently underestimates muscle mass compared to DEXA measurements. While this systematic underestimation represents a predictable offset that could potentially be addressed through calibration protocols, it is an important consideration for diagnostic implementation. The 95% limits of agreement ranging from -8.158 kg to $+2.851$ kg represent the expected range of measurement differences between the two methods for individual patients.

For diagnostic applications, this measurement variability has important clinical implications, particularly for patients near sarcopenia diagnostic thresholds where differences within this range could influence classification decisions. However, the systematic nature of the bias suggests that with appropriate calibration and establishment of method-specific diagnostic cutoffs, our approach could serve as a reliable diagnostic tool rather than merely a screening instrument. The predictable offset indicates that the measurement differences are not random but follow a consistent pattern, which is advantageous for developing standardized diagnostic protocols. With further refinement and validation studies to establish optimal diagnostic thresholds specific to X-ray-derived measurements, this approach demonstrates significant potential for accurate sarcopenia diagnosis, particularly in clinical scenarios where DEXA limitations compromise diagnostic accuracy.

4.3. Advantages in Implant Patients

While DEXA is widely used for measuring bone mineral density and body composition, including muscle mass, the presence of implants can introduce significant inaccuracies. Metal implants, such as those used in joint replacements, can lead to overestimation of bone mineral content (BMC) and soft-tissue mass. Studies have shown that metal rods can inflate whole-body BMC measurements by 1.5–3% and significantly increase reported soft-tissue mass ($p < 0.003$) [26]. Similarly, breast implants have been found to increase whole-body BMC by 4.7% [27]. These inaccuracies extend to muscle mass measurements, with total knee replacements (TKR) causing overestimation of lean mass in affected limbs [28]. To address these issues, correction methods such as Automatic Metal Detection (AMD) processing and substitution protocols have been developed [29]. AMD processing has been shown to mitigate overestimation of muscle mass in patients with TKR, while substitution protocols involving the replacement of affected regions with measurements from unaffected contralateral sides have improved the accuracy of body composition assessments. These findings underscore the importance of accounting for implants when interpreting DEXA results, particularly in conditions like sarcopenia where precise muscle mass measurement is crucial. Clinicians must be aware of these potential inaccuracies to avoid misdiagnosis and ensure appropriate treatment decisions.

However, our AI-based muscle segmentation model using long extremity X-rays presents a promising solution to the limitations of DEXA in patients with implants. By directly analyzing image features to delineate muscle boundaries, rather than relying on X-ray attenuation, the AI model can potentially mitigate implant-related artifacts and provide more accurate muscle mass estimations. The model's ability to adapt through learning, focus on specific muscle groups, and operate independently of density measurements offers advantages over DEXA in complex anatomical scenarios. This approach could lead to more consistent and reliable muscle mass quantification across diverse patient popula-

tions, including those with prosthetics or joint replacements. While further validation is necessary, particularly in patients with various implant types, our AI model shows potential for improving the accuracy of muscle mass assessment and longitudinal tracking in patients with implants, potentially enhancing the diagnosis and monitoring of conditions like sarcopenia.

One case in the cohort illustrates the limitations of using DEXA to diagnose sarcopenia in patients who have implants. The patient was diagnosed with an intertrochanteric fracture and underwent intramedullary surgery. DEXA scans were performed immediately after surgery and at a one-year follow-up. As Figure 5 shows, the one-year follow-up DEXA scan indicated a significant increase in the skeletal muscle index (from 5.15 to 6.31 kg/m², 22.5% increased), incorrectly suggesting an improvement from severe sarcopenia to not severe sarcopenia. However, gait analysis results contradicted the DEXA findings and lower extremity X-rays revealed significant muscle area loss (from 77 to 71.5 cm², 7.15% decreased) in one-year follow-up X-ray.

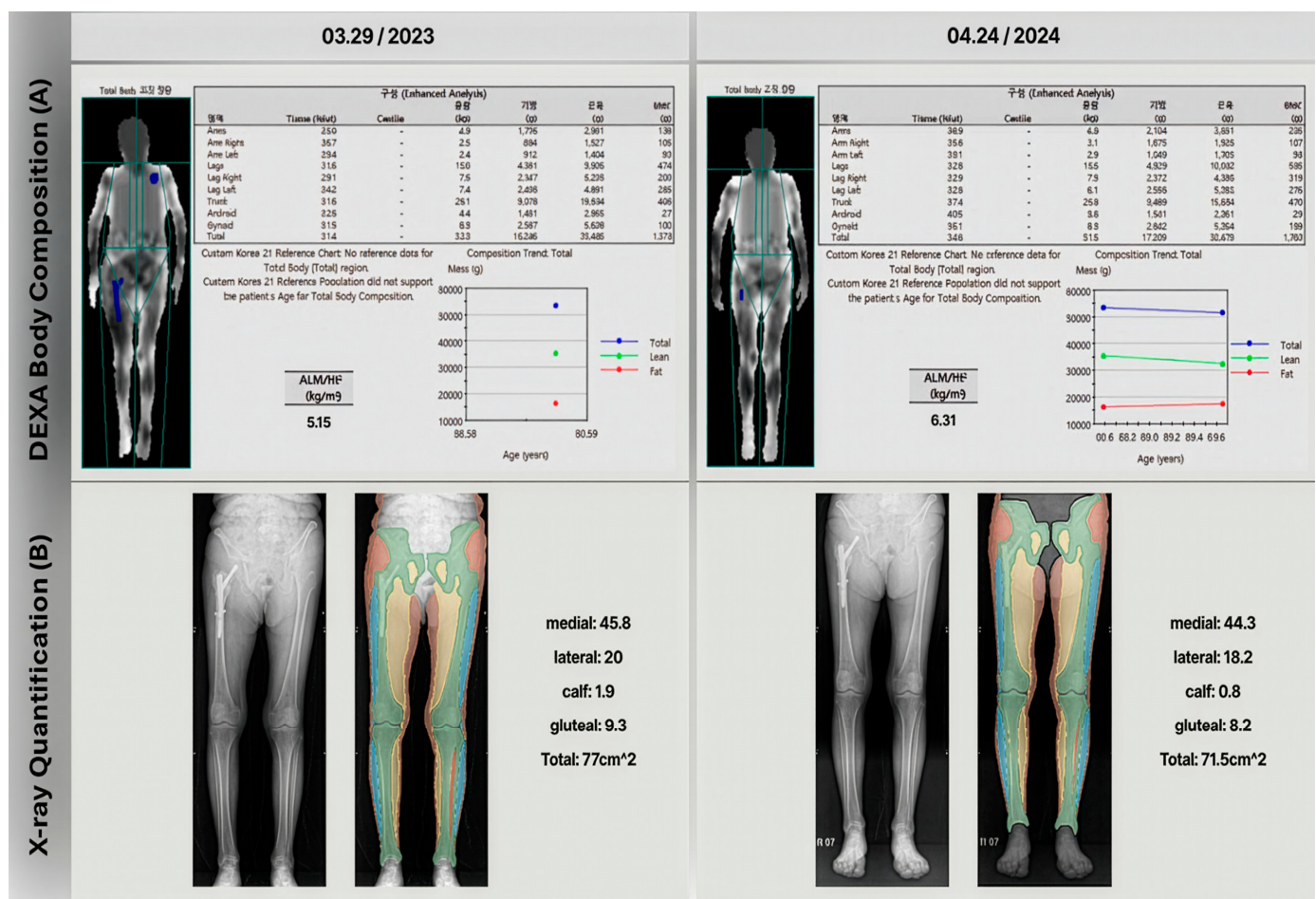


Figure 5. A single patient's muscle measurements are shown at two time points (29 March 2023 and 24 April 2023). (A) displays DEXA body composition results with Korean column headers indicating "구성 (Composition)", "영역 (Region)", "총량 (Total mass)", "지방 (Fat mass)", and "근육 (Lean mass)", while (B) presents corresponding X-ray-based muscle region quantifications. The color-coded segments illustrate changes in medial, lateral, and gluteal muscle areas over time.

This case is particularly relevant because patients with the sarcopenia range are at higher risk for fractures and knee osteoarthritis, often requiring arthroplasty. These findings suggest that muscle estimation methods based on lower extremity X-ray segmentation method could be a viable alternative to DEXA for sarcopenia assessment in such cases. This

approach may provide more accurate results, especially in patients who have undergone orthopedic surgeries that could potentially affect DEXA measurements.

4.4. Limitations and Future Directions

Our study has several limitations that provide clear directions for future research. First, there is significant selection bias due to the single-center design, as data was collected solely from a university medical center in South Korea with a cohort primarily composed of patients with hip fractures or hip-related orthopedic conditions. To address this limitation and enhance generalizability, multi-center studies incorporating diverse populations, age ranges, and various musculoskeletal conditions beyond hip fractures are essential for broader clinical applicability.

Second, while our model demonstrates statistical correlation with DEXA measurements ($r = 0.72$), the clinical significance and diagnostic implications require more robust evaluation, particularly given the wide limits of agreement (95% limits: -8.158 kg to $+2.851\text{ kg}$) that may impact diagnostic decisions near sarcopenia thresholds. Future studies should conduct comprehensive diagnostic accuracy analyses, including sensitivity and specificity assessments for sarcopenia diagnosis, and investigate the impact on clinical decision-making and patient outcomes.

Third, our model focuses on muscle quantity rather than quality or function, which are crucial factors in comprehensive sarcopenia assessment. Integration of functional data such as gait analysis, grip strength, and Short Physical Performance Battery (SPPB) scores alongside muscle quantity measurements will provide more comprehensive musculoskeletal health evaluation.

Finally, our cross-sectional design and primary comparison with DEXA limit validation scope. Future research should include longitudinal studies to track muscle mass changes over time and comparisons with other imaging modalities like CT and MRI. These developments will enable broader clinical applications including early sarcopenia screening, treatment monitoring in rehabilitation programs, and personalized intervention strategies across diverse patient populations in musculoskeletal medicine.

5. Conclusions

Our AI-based semantic segmentation model achieved excellent performance (IoU = 0.93, Dice = 0.95) and strong correlation with DEXA ($r = 0.72$, $p < 0.0001$) in quantifying muscle mass from lower extremity X-rays. This approach offers significant advantages over DEXA, particularly in patients with artificial implants where conventional methods can lead to sarcopenia misdiagnosis. The model provides detailed muscle group information while mitigating implant-related measurement errors.

Despite limitations including single-center design, the need for clinical significance validation, and functional assessment integration, this study represents a significant advancement in muscle mass quantification. Future research should conduct multi-center validation studies, establish diagnostic accuracy metrics, and integrate functional assessments through longitudinal studies. This AI-driven approach has the potential to revolutionize sarcopenia diagnosis and monitoring, especially in complex patient populations where current methods fail.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/healthcare13192488/s1>, Figure S1: Result of semantic segmentation model.

Author Contributions: Conceptualization, H.P. and J.Y.; Methodology, H.K.; Software, H.K.; Validation, H.P., H.K. and J.Y.; Investigation, H.K.; Data Curation, H.P. and H.K.; Writing—Original Draft Preparation, H.P. and H.K.; Writing—Review and Editing, H.P. and J.Y.; Visualization, H.K.; Supervision, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Health & Welfare, Republic of Korea (Grant No. RS-2024-00507256), and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant No. RS-2021-NR060097).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (IRB No. INUH 2023-04-027) on 28 July 2024 at Inha University Hospital.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alghannam, A.F.; Alharbi, D.S.; Al-Hazzaa, H.M. Sarcopenia of Ageing: Does a Healthier Lifestyle Matter in Reversing the Trajectory? A Brief Narrative Review and a Call for Action in Saudi Arabia. *Saudi J. Med. Med. Sci.* **2024**, *12*, 10–16. [\[CrossRef\]](#)
2. Papadopoulou, S.K. Sarcopenia: A Contemporary Health Problem among Older Adult Populations. *Nutrients* **2020**, *12*, 1293. [\[CrossRef\]](#)
3. Larsson, L.; Degens, H.; Li, M.; Salvati, L.; Lee, Y.I.; Thompson, W.; Kirkland, J.L.; Sandri, M. Sarcopenia: Aging-Related Loss of Muscle Mass and Function. *Physiol. Rev.* **2019**, *99*, 427–511. [\[CrossRef\]](#)
4. Petermann-Rocha, F.; Balntzi, V.; Gray, S.R.; Lara, J.; Ho, F.K.; Pell, J.P.; Celis-Morales, C. Global prevalence of sarcopenia and severe sarcopenia: A systematic review and meta-analysis. *J. Cachexia Sarcopenia Muscle* **2021**, *13*, 86–99. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Wang, Y.; Wei, X. Sarcopenia, the Economic Challenges in Healthcare and Individual Struggles Arise. *OALib* **2024**, *11*, 1–7. [\[CrossRef\]](#)
6. Kim, D.; Oh, K. Prevalence of Sarcopenia in the Republic of Korea. *PHWR* **2024**, *17*, 1055–1067.
7. Swan, L.; Warters, A.; O’Sullivan, M. Socioeconomic Disadvantage is Associated with Probable Sarcopenia in Community-Dwelling Older Adults: Findings from the English Longitudinal Study of Ageing. *J. Frailty Aging* **2022**, *11*, 398–406. [\[CrossRef\]](#)
8. Ben Kirk, B.; Cawthon, P.M.; Arai, H.; Ávila-Funes, J.A.; Barazzoni, R.; Bhasin, S.; Binder, E.F.; Bruyere, O.; Cederholm, T.; Chen, L.-K.; et al. The Conceptual Definition of Sarcopenia: Delphi Consensus from the Global Leadership Initiative in Sarcopenia (GLIS). *Age Ageing* **2024**, *53*, afae052. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Ooi, H.; Welch, C. Obstacles to the Early Diagnosis and Management of Sarcopenia: Current Perspectives. *Clin. Interv. Aging* **2024**, *19*, 323–332. [\[CrossRef\]](#)
10. Park, H.; Kim, H.S.; Gu, B.-S.; Kim, H.; Yoo, J.-I. Latest Updates on Sarcopenia and Cachexia: Insights from the 17th Sarcopenia, Cachexia, and Wasting Disorders Conference. *J. Bone Metab.* **2025**, *32*, 167–179. [\[CrossRef\]](#)
11. Cho, M.-R.; Lee, S.; Song, S.-K. A Review of Sarcopenia Pathophysiology, Diagnosis, Treatment and Future Direction. *J. Korean Med. Sci.* **2022**, *37*, e146. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Chen, L.-K.; Woo, J.; Assantachai, P.; Auyeung, T.-W.; Chou, M.-Y.; Iijima, K.; Jang, H.C.; Kang, L.; Kim, M.; Kim, S.; et al. Asian Working Group for Sarcopenia: 2019 Consensus Update on Sarcopenia Diagnosis and Treatment. *J. Am. Med. Dir. Assoc.* **2020**, *21*, 300–307.e302. [\[CrossRef\]](#)
13. Cruz-Jentoft, A.J.; Bahat, G.; Bauer, J.; Boirie, Y.; Bruyère, O.; Cederholm, T.; Cooper, C.; Landi, F.; Rolland, Y.; Sayer, A.A.; et al. Sarcopenia: Revised European consensus on definition and diagnosis. *Age Ageing* **2019**, *48*, 16–31. [\[CrossRef\]](#)
14. Messina, C.; Albano, D.; Gitto, S.; Tofanelli, L.; Bazzocchi, A.; Ulivieri, F.M.; Guglielmi, G.; Sconfienza, L.M. Body composition with dual energy X-ray absorptiometry: From basics to new tools. *Quant. Imaging Med. Surg.* **2020**, *10*, 1687–1698. [\[CrossRef\]](#)
15. Kim, D.W.; Ha, J.; Ko, Y.; Kim, K.W.; Park, T.; Lee, J.; You, M.-W.; Yoon, K.-H.; Park, J.Y.; Kee, Y.J.; et al. Reliability of Skeletal Muscle Area Measurement on CT with Different Parameters: A Phantom Study. *Korean J. Radiol.* **2021**, *22*, 624–633. [\[CrossRef\]](#)
16. Faron, A.; Sprinkart, A.M.; Kuetting, D.L.R.; Feisst, A.; Isaak, A.; Endler, C.; Chang, J.; Nowak, S.; Block, W.; Thomas, D.; et al. Body composition analysis using CT and MRI: Intra-individual intermodal comparison of muscle mass and myosteatosis. *Sci. Rep.* **2020**, *10*, 11765. [\[CrossRef\]](#)

17. Lee, S.Y.; Ahn, S.; Kim, Y.J.; Ji, M.J.; Kim, K.M.; Choi, S.H.; Jang, H.C.; Lim, S. Comparison between Dual-Energy X-ray Absorptiometry and Bioelectrical Impedance Analyses for Accuracy in Measuring Whole Body Muscle Mass and Appendicular Skeletal Muscle Mass. *Nutrients* **2018**, *10*, 738. [[CrossRef](#)] [[PubMed](#)]
18. Coppini, L.Z.; Waitzberg, D.L.; Campos, A.C.L. Limitations and validation of bioelectrical impedance analysis in morbidly obese patients. *Curr. Opin. Clin. Nutr. Metab. Care* **2005**, *8*, 329–332. [[CrossRef](#)]
19. Sato, H.; Nakamura, T.; Kusuhashi, T.; Kenichi, K.; Kuniyasu, K.; Kawashima, T.; Hanayama, K. Effectiveness of impedance parameters for muscle quality evaluation in healthy men. *J. Physiol. Sci.* **2020**, *70*, 53–59. [[CrossRef](#)]
20. Urooj, B.; Ko, Y.; Na, S.; Kim, I.-O.; Lee, E.-H.; Cho, S.; Jeong, H.; Khang, S.; Lee, J.; Kim, K.W. Implementation of Fully Automated AI-Integrated System for Body Composition Assessment on Computed Tomography for Opportunistic Sarcopenia Screening: Multicenter Prospective Study. *JMIR Form. Res.* **2025**, *9*, e69940. [[CrossRef](#)] [[PubMed](#)]
21. Kim, H.S.; Kim, H.; Kim, S.; Cha, Y.; Kim, J.-T.; Kim, J.-W.; Ha, Y.-C. Precise individual muscle segmentation in whole thigh CT scans for sarcopenia assessment using U-net transformer. *Sci. Rep.* **2024**, *14*, 3301. [[CrossRef](#)] [[PubMed](#)]
22. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Volume 9351, pp. 234–241. [[CrossRef](#)]
23. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [[CrossRef](#)]
24. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: 4th International Workshop on Deep Learning in Medical Image Analysis, DLMIA 2018 and 8th International Workshop on Multimodal Learning for Clinical Decision Support, ML-CDS 2018 Held in Conjunction with MICCAI 2018. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support—4th International Workshop, DLMIA 2018 and 8th International Workshop, ML-CDS 2018 Held in Conjunction with MICCAI 2018*; Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.M.R.S., Bradley, A., Papa, J.P., Belagiannis, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11. [[CrossRef](#)]
25. Hwang, D.; Ahn, S.; Park, Y.-B.; Kim, S.H.; Han, H.-S.; Lee, M.C.; Ro, D.H. Deep Learning-Based Muscle Segmentation and Quantification of Full-Leg Plain Radiograph for Sarcopenia Screening in Patients Undergoing Total Knee Arthroplasty. *J. Clin. Med.* **2022**, *11*, 3612. [[CrossRef](#)]
26. Giangregorio, L.M.; Webber, C.E. Effects of metal implants on whole-body dual-energy x-ray absorptiometry measurements of bone mineral content and body composition. *Can. Assoc. Radiol. J.* **2003**, *54*, 305.
27. Donlon, C.M.; Chou, S.H.; Yu, C.Y.; LeBoff, M.S. Accounting for Surgical Confounding Factors Affecting Dual-Energy X-ray Absorptiometry in a Large Clinical Trial. *J. Clin. Densitom.* **2021**, *25*, 127–132. [[CrossRef](#)]
28. Jang, J.Y.; Kim, M.; Lee, D.; Won, C.W. Effect of total knee replacement on skeletal muscle mass measurements using dual energy X-ray absorptiometry. *Sci. Rep.* **2023**, *13*, 2908. [[CrossRef](#)] [[PubMed](#)]
29. Harper, K.D.; Clyburn, T.A.; Incavo, S.J.; Lambert, B.S. DEXA overestimates bone mineral density in adults with knee replacements. *Sports Med. Health Sci.* **2020**, *2*, 211–215. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.