Zhang et al, Gerontechnology (2026) 25:4
https://doi.org/10.30289/gt.202606_25(4).1703

Automating multisite muscle segmentation in ultrasound via deep learning through a standardized framework for sarcopenia assessment

Dawei Zhang MSc[a], Chonglin Wu BSc[a], Yuxuan Na MSc[a], Wanrui Li MSc[a], Ka-Shing Lee MSc[a], Yu Sun MD PhD[a], Yongping Zheng PhD[a,b]*

[a]Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China; [b]Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hong Kong SAR, China
**Corresponding author: yongping.zheng@polyu.edu.hk*

Abstract
Background: Nowadays, sarcopenia is widely affecting numerous people in the world, which directly leads to devastating clinical outcomes like catastrophic falls, functional disability in mobility, and increased mortality. Ultrasound (US) plays an important role in sarcopenia assessment for quantitatively and qualitatively detecting muscle performance by muscle thickness (MT) and cross-sectional area (CSA). However, it has been hindered by operator dependency and a lack of standardization, making manual measurements impractical for large-scale screening. Research Aims Our study addresses this clinical bottleneck by adopting image segmentation models to enable deep-learning-based automated ultrasound system that standardizes the assessment of multiple muscle groups across the body.
Methods: We trained the deep learning models (U-Net and nnU-Net) for segmenting five key muscle groups (biceps, triceps, rectus abdominis (RA), rectus femoris(RF), and peroneal longus and brevis (PLPB) from US images acquired from 94 adult participants (young (n=22), middle-aged (n=22), and elder (n=50)). The models' segmentation performance was evaluated against expert manual annotations using the Dice Similarity Coefficient (Dice). For the predicted dataset, the results are correlated with body composition results and physical performance data.
Results: The deep learning segmentation achieved excellent performance across all muscle groups. The mean Dice were: RF (0.8260), RA (0.8810), biceps (0.8370), triceps (0.8080), and PLPB (0.9200), indicating a high accuracy compared to the ground truth. Furthermore, these AI-derived metrics demonstrated strong clinical validity through high correlations with body composition data. The appendicular muscles has the highest associations found between triceps CSA and MM (r=0.8212), PLPB MT and ASM (r=0.7462), and RA CSA and BMR (r=0.6398).
Conclusions: By applying a deep learning-powered framework, the ultrasound provides accurate, highly efficient segmentation of multiple muscles for sarcopenia assessment, making US a practical method for sarcopenia assessment and enabling its adoption in community-based screening and the longitudinal management of sarcopenia.

Keywords: ultrasound, sarcopenia, deep learning, muscle segmentation, muscle thickness

Introduction

Sarcopenia, the progressive and generalized loss of skeletal muscle mass and function, represents a major public health concern in aging societies worldwide (Cho et al., 2022). Formally recognized as a muscle disease with its own ICD-10 code since 2016, sarcopenia is directly associated with a cascade of devastating clinical outcomes, including catastrophic falls, functional disability, and increased mortality (Cao & Morley, 2016). While interventions such as resistance training and adequate protein intake can mitigate its progression, their effectiveness hinges on early and accurate detection [4]. This places a critical emphasis on developing accessible, reliable, and quantitative assessment tools to identify at-risk individuals before irreversible functional decline occurs.

Asian Working Group for Sarcopenia (AWGS) and European Working Groups for Sarcopenia (EWGSOP) provided consensus on the sarcopenia screening and diagnosis(Chen et al., 2014; Chen et al., 2020; Cruz-Jentoft et al., 2010; Cruz-Jentoft et al., 2019a). However, most modern technologies focus on the use of Bioelectrical Impedance Analysis (BIA) and Dual X-ray Absorptiometry (DXA) to evaluate appendicular skeletal muscle mass (ASM) (Guglielmi et al., 2016; Tagliafico et al., 2022). Other diagnosis methods include mostly the physical perfor-

mance, such as the time-up-and-go (TUG) and hand gripping tests. Other than the BIA and DXA, ultrasound has been drawing more and more attention to modern time sarcopenia evaluation due to its economic efficacy, portability, and accuracy in measurement(Chen et al., 2022; Hida et al., 2018). BIA and DXA lack accuracy in some cases, especially for patients who are measured after urination and exercise. The total body water (TBW) could be underestimated and overestimated, as certain parameters, such as the fat-free mass (FFM) and ASM in the body, are involved, thus providing inaccurate results. However, ultrasound could detect muscle morphology at any time, unaffected by exercise or metabolic activity (Chen et al., 2025).

In the consensus of the EWGSOP, ultrasound has been recommended for sarcopenia bedside evaluation, especially in community and in-home care settings(Cruz-Jentoft et al., 2019a, 2019b). However, ultrasound requires a large number of professional doctors and practitioners for image diagnosis and reading(Abe et al., 2018). Ultrasound used in the clinical setting for sarcopenia has many bottlenecks to overcome: the operator-dependency, subjectivity in the interpretation, and the inefficiency (Behboodi et al., 2024; Guo et al., 2011). Firstly, different operators could collect different images; secondly, different operators could interpret the muscles differently, such as reading the borders; thirdly, the manual labelling of the muscle is very time-consuming, and sarcopenia requires the multi-site muscles' overall interpretation, especially for a large-scale data collection in the community, where manual measurement is very impractical. Therefore, to deploy an AI-powered, automated framework for multisite muscle segmentation from ultrasound images could hugely benefit screening process (Yi et al., 2022). Our primary goal is to overcome the limitations of manual ultrasound analysis, thereby creating a standardized, efficient, and objective tool for sarcopenia assessment, especially using the methods of deep learning to improve the efficiency of image segmentation and provide real-time segmentation of the muscles (Behboodi et al., 2024; Zhang et al., 2024).

Major muscle groups can indicate how the muscle behaves through its thickness and cross-sectional area. In most cases, the rectus femoris (RF) is one of the most renowned muscle groups of the thigh that helps assess elderly people's movements and whether they are at risk of falling (Chen et al., 2022; Nies et al., 2022). The calf is also important to reflect the mobility of elderly people. The gait and the falling in strongly correlated with the peroneus longus and brevis (PLPB) (Kuyumcu et al., 2016; Rolland et al., 2003). Therefore, the PLPB assessment is important for assessing how inclined elderly people are to fall based on PLPB quality and quantity, though the thigh is more affected by sarcopenia (Fuchs et al., 2023). The biceps and triceps could indicate the performance of the muscles in the arm, which elderly people might feel strengthless in weightlifting due to sarcopenia(Chang et al., 2018; Li et al., 2020). Under the consensus made by the AWGS and EWGSOP, handgrip is taken for the upper limbs' capacity exam; however, a straightforward upper limbs muscle assessment has not been added to the norm of sarcopenia evaluation. Thus, the biceps and triceps are important in the muscle segmentation for sarcopenia evaluation(Larsson et al., 2019). By measuring the muscles from all the above, a multi-site, global muscle screening using ultrasound could be conducted to determine the severity of sarcopenia on a muscle-by-muscle basis, using measures such as muscle thickness, cross-sectional area, and volume.

In our studies, we aimed to use deep learning methods on various ultrasound-detected parameters for sarcopenia evaluation and conquer the bottlenecks that ultrasound presents, especially in clinical settings: subjectivity, low efficiency, and operator dependencies. We have obtained major muscle groups from various elderly patients and segmented the major muscles. To achieve the best segmentation outcome, we utilized the deep learning convolutional structure of U-Net and nnU-Net. With this quick and real-time analysis, ultrasound could be used in a more efficient way so that sarcopenia could be easily measured with higher accuracy. However, the restriction of ultrasound assessment to specialized radiology departments creates a significant bottleneck for large-scale geriatric care. To bridge the gap between advanced medical imaging and accessible community health services, this study aligns with the core principles of gerontechnology by developing a tool specifically designed for non-specialist use. Our goal is not merely to improve image reading efficiency, but to transform ultrasound into a practical, "point-of-care" screening device for sarcopenia as well. By automating the expert interpretation process, we aim to empower geriatricians, nurses, and community health workers to objectively assess muscle quality in aging populations, thereby facilitating early intervention and supporting aging-in-place strategies.

Methodology

This study was based on ultrasound data prospectively collected from 94 adult participants. The selecting criteria includes elderly people

older than 60 years old and capable of physical performance of walking and have no contraindications for ultrasound scanning. The data was used to develop and validate a deep learning framework for the automated segmentation of five key muscle groups RF, RA, biceps, triceps, and PLPB. The deep learning framework of U-Net and nnU-Net was employed for the segmentation task. A standard U-Net was used for the RF, RA, biceps, and triceps, as these typically appear as single, well-defined structures in static images. For the more challenging task of accurately differentiating the adjacent peroneal longus and brevis muscles, the self-configuring nnU-Net framework was utilized, given its superior capability in handling complex, multi-class segmentation problems.

The study received ethical approval from the Human Subject Ethics Sub-committee (HSESC) of The Hong Kong Polytechnic University (Ref: HSEARS 20250528002, HSEARS 20250610004), and all participants provided written informed consent. All ultrasound scans were performed by two trained operators under the supervision of a physiatrist with over ten years of experience in musculoskeletal ultrasound. A high-frequency linear transducer (128 elements, 7.5/10 MHz) was used for all acquisitions, with the imaging depth set to 40 mm. The BIA used is from the Tanita DC-430MA (Tanita Health Equipment Hong Kong). To ensure consistency, a standardized protocol was strictly followed. At each of the five muscle sites, the transducer was oriented transversely to the muscle fibers to capture a clear cross-sectional view. A water-soluble gel (Aquasonic Ultrasound Gel, Parker Laboratories, Fairfield, NJ, USA) was applied with moderate pressure to avoid muscle compression. For each site, two images were captured, and the one with the clearest fascial borders was selected for analysis and exported as a 384x400 pixel grayscale file.

Figure 1 provides a schematic overview of the entire study workflow. The process began with subject recruitment and the acquisition of primary clinical data, including BIA and functional performance tests (Hand-gripping, Time-Up-and-Go). Concurrently, ultrasound images were systematically collected from five key muscle groups. The entire image dataset was then partitioned at the subject based that split into training (70%), validation (10%), and testing (20%) sets. An experienced ultrasound doctor supervised the manual annotation process using LabelMe to generate the ground truth segmentation masks. This dataset was used to train and validate the U-Net and nnU-Net model. Finally, the performance of the trained models was evaluated by comparing the automated segmentation results on the unseen test set with the expert-annotated ground truth.

To ensure standardized and clinically meaningful quantitative analysis, a calibration process was implemented to convert pixel-based measurements to metric units (millimetres). Due to variations in the acquisition depth across the ultrasound images, direct pixel measurements are not comparable. Therefore, for each image, a unique spatial resolution factor, or pixel spacing (mm/pixel), was calculated by dividing the known real-world scanning depth (in mm) by the image's height in pixels. This calibration factor was subsequently used to convert all measurements: muscle thickness was determined by multiplying the pixel thickness by the pixel spacing, and the CSA was calculated by multiplying the total pixel count of the segmented region by the square of the pixel spacing. This procedure guaranteed that all derived anatomical measurements were transformed to a consistent physical scale, enabling valid and comparable analysis.

2.1 Scanning protocols

2.1.1 Rectus Femoris (RF)

For the RF, RA, biceps, and triceps, a static acquisition technique was used, where the optimal cross-sectional view was identified and captured at a single, predefined anatomical landmark. In contrast, for the complex peroneal muscle group of PLPB, a dynamic scanning technique was employed. The specific scanning protocol for the RF involved having the subject sit in a chair with their feet flat on the floor. The thigh is parallel to the ground, and the RF muscle is entirely shown on the screen. The scanning begins from the knee and moves 10-15 centimetres to the RF muscle. Our imaging protocol incorporated two distinct acquisition modes tailored to the anatomy of the target muscles (*Figure 2* (1)).

2.1.2 Rectus Abdominals (RA)

With the participant in a supine position, the transducer was initially placed over the umbilicus to identify the linea alba (the connective tissue midline). The transducer was then moved approximately 3 cm laterally to the right side, ensuring the medial and lateral fascial borders of the RA muscle were clearly visualized for measurement (*Figure 2* (2)).

2.1.3 Biceps

Participants were in a supine position with their arm resting alongside the body in a neutral position, palm facing up. The elbow was fully extended, and the muscle was relaxed. The probe was placed transversely at the middle point of the distance from the acromion process to the

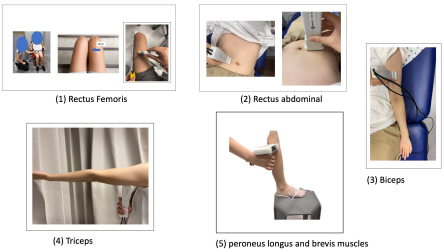

Figure 1. The process for the multi-site scanning for the sarcopenia assessment via ultrasound

cubital fossa, and different subjects could vary the position, upper or lower, due to individual dependencies (*Figure 2* (3)).

2.1.4 Triceps

Participants were standing, placing their hand on a countertop at shoulder height on the opposite side to stabilize the body. The measured arm was relaxed, hanging naturally at the side. The probe was placed transversely on the posterior aspect of the arm at the midpoint between the acromion process and the olecranon (*Figure 2*(4)).

2.1.5 Peroneus Brevis and Longus

Participants were standing, with the foot of the leg being measured placed on a stool at approximately waist height, creating a flexed position at the hip and knee. The operator first identified the muscles distally near the ankle and scanned proximally up the lateral aspect of the lower leg. The target image was captured at the proximal third of the leg, where the muscle belly of the peroneus longus was maximally visible before tapering into its tendon near the knee (*Figure 2* (5)). Specifically, the ultrasound image is taken using the Scolioscan Air Probe® (Zheng et al., 2016) due to the larger coverage of the probe for the calf muscles from a transverse panel.

2.2 Deep learning framework for automated segmentation

We used automated segmentation methods based on U-Net and nnU-Net. The image dataset, comprising images from all 94 subjects, was partitioned at the subject level to prevent data leakage. Specifically, subjects were randomly assigned to a training set (70%, n=66), a validation set (10%, n=9), and a test set (20%, n=19). In order to train the data, we have manually labelled the images via LabelMe (Version 5.5.0). The results are shown in *Figure 3*, which presents examples of segmentation from the RF, RA, Biceps, Triceps, and PLPB.

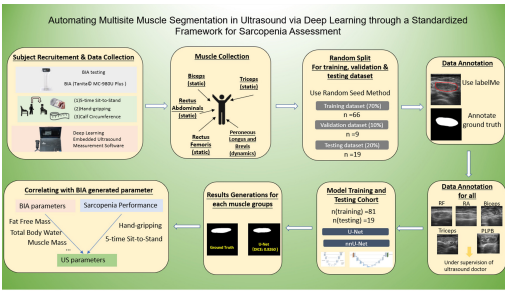

Figure 2. Scanning protocol example of the muscles

For the segmentation of the RF, RA, biceps, and triceps, a standard U-Net architecture was implemented (Ronneberger et al., 2015). The U-Net's encoder-decoder structure with skip connections is highly effective for capturing both high-level contextual information and fine-grained spatial details, making it ideal for precise localization of muscle boundaries. As a widely adopted convolutional neural network (CNN) in medical image analysis, U-Net features a symmetric encoder-decoder structure. The contracting path (encoder) captures high-level semantic context through successive convolutional layers and max-pooling operations, effectively reducing spatial resolution while extracting features. Conversely, the expansive path (decoder) recovers spatial information and precise localization by up-sampling the feature maps. Crucially, U-Net employs skip connections that concatenate feature maps from the encoder directly to the corresponding layers in the decoder. This mechanism preserves fine-grained spatial details lost during pooling, enabling the network to generate highly accurate segmentation masks even with limited training data, which is characteristic of medical ultrasound datasets.

For the more challenging task of differentiating the adjacent peroneus longus and brevis muscles, the nnU-Net framework was employed (Isensee et al., 2018). U-Net means no-new U-Net, which is a structure built on the U-Net architecture. As an adaptive, self-configuring framework, nnU-Net automatically optimizes the network architecture, pre-processing steps, and training scheme based on the specific characteristics of the input peroneal dataset, thereby maximizing segmentation performance without manual parameter tuning.

In contrast, the PLPB group presented a significantly more complex segmentation challenge, necessitating the use of the nnU-Net framework. Anatomically, the PLPB are adjacent muscles that often exhibit similar echogenicity, making the inter-muscular septum difficult to distinguish. Furthermore, the PLPB dataset was derived from dynamic video frames, introducing greater variability in muscle shape, probe incidence angles, and speckle noise than static acquisitions at other sites. U-Net is a self-configuring framework that

addresses these challenges by automatically adapting the network topology and pre-processing steps such as normalization and resampling. This adaptive capability allows nnU-Net to better handle the subtle textural differences and high variability of the PLPB data, ensuring superior segmentation performance where a standard, manually tuned U-Net might struggle to generalize.

All models were trained using a combined Dice and cross-entropy loss function with an Adam optimizer. The initial learning rate was set to 1e-4, and the batch size was 8. To enhance model robustness and mitigate overfitting, on-the-fly data augmentation techniques were applied to the training set, including random rotations, scaling, and elastic deformations. Training was monitored using the validation set, and the model weights that achieved the highest Dice score on the validation data were saved for final testing. Early stopping was implemented if the validation performance did not improve for 20 consecutive epochs.

2.3 Performance evaluation and statistical analysis

The performance of the final trained models was rigorously evaluated on the test set (n=19 subjects). To provide a comprehensive assessment of segmentation accuracy, we calculated two key region-based metrics by comparing the AI-generated masks against the expert's ground truth delineations. The first is to use the dice similarity coefficient (Dice). The Dice is the primary metric, measuring the volumetric overlap between the predicted and ground truth masks. It is calculated as Equation 1, where A and B are the predicted and ground truth masks, respectively. The second evaluation method is to use Intersection Over Union (IoU). This metric measures the extent of overlap and is calculated as shown in Equation 2. The third evaluation metric is to use 95th Percentile Hausdorff Distance (HD95) (Equation 3). HD could tell the boundary-based metric that measures the maximum distance between the predicted and ground truth boundaries. By using the 95th percentile, it offers robustness against a small number of outliers, providing a more stable assessment of boundary deviation, measured in pixels.

$$IoU = \frac{|A \cap B|}{|A \cup \mathrm{B}|}$$

Equation 1. Dice Similarity Coefficient (DICE) Equation

$$dice = \frac{2 \times A \cap B}{|A| + |B|}$$

Equation 2. Intersection Over Union (IoU) Equation

$$H\ (A, B) = \max\{\sup a \in A \inf b \in B\ ||a - b||, \sup b \in B \inf a \in A||a - b||\}$$

Equation 3. Hausdorff Distance (HD95) Equation

Results

3.1 Participant characteristics

A total of 94 participants were included in this study, with an overall mean age of 52.0 ± 19.3 years and 64 (68.1%) being female. Specifically, for the PLPB scanning using Scolioscan Air, the study included 12 subjects, resulting in 3465 images. The cohort was stratified into three groups based on age: Young (n = 22), Middle-aged (n = 22), and Older (n = 50). The detailed baseline demographics and clinical characteristics for the overall cohort (for RF, RA, and thigh) and for each subgroup are presented in *Table 1*.

Given the unique anatomical complexity of the PLPB, a specialized data strategy was employed to ensure model robustness and generalizability. A dedicated, multi-source dataset was constructed specifically for this task, separate from the other four muscle groups. The training and validation set comprised a total of 46 subjects. This dataset was collected at our facility using the Scolioscan Air system. Within the dataset, 12 subjects are categorized into the testing set. These subjects were entirely unseen by the model during training and validation, providing a rigorous test of the model's real-world generalization capability.

3.2 segmentation results and comparisons of segmentations with the annotations

The deep learning models achieved excellent segmentation accuracy across all five evaluated muscle groups when compared to the expert's manual annotations. *Table 2* provides a detailed summary of the performance metrics on the unseen test set. The mean Dice was consistently high, ranging from 0.808 for the triceps to 0.9200 for the peroneal muscles. Specifically, the mean (± standard deviation) DSC values were: RF (0.8260 ± 0.07), RA (0.8810 ± 0.05), biceps (0.8370 ± 0.08), triceps (0.8080 ± 0.09), and the peroneus longus and brevis group (0.9200 ± 0.04). It is worth noting that PLPB is obtained with videos, which are then turned into frames, providing more images than other muscle groups. *Figure 3* provides visual examples of the automated segmentation results overlaid on the original ultrasound images, alongside the manual ground truth. These qualitative examples demonstrate a high degree of spatial overlap between the predicted masks and the ground truth annotations.

Table 1. Demographics and Subject Characteristics (RF, RA, Biceps and Triceps)

Characteristics	Young (n=22)	Middle-aged(n=22)	Elder (n=50)	Overall(n=94)
Age (years)	24.5 ± 2.1	51.2 ± 5.6	76.8 ± 6.5	52.0 ± 19.3
Gender (Female, n (%))	16(72.7%)	14(63.6%)	34(68.0%)	64(68.1%)
Body Mass Index (kg/m²)	21.8 ± 2.5	24.5 ± 3.1	23.1 ± 2.8	23.1 ± 2.9
Waist Circumference (cm)	76.3 ± 8.1	85.1 ± 9.5	81.5 ± 9.0	80.9 ± 9.3
Handgrip Strength (kg)	35.2 ± 6.8	28.5 ± 5.9	21.3 ± 5.1	26.5 ± 8.2
Five Times Sit-to-Stand (s)	8.9 ± 1.5	11.2 ± 2.1	14.5 ± 3.0	12.3 ± 3.4

3.3 Results of predicted outcome and the ground truth

In order to compare the results from the predicted segmentation with the ground truth, the quantitative accuracy of the segmentation was assessed by comparing the predicted muscle thickness against ground truth annotations using methods of correlation and Bland-Altman analysis for the RF, RA, PLPB, biceps and triceps.

For the RF muscle of the thigh, our model demonstrated strong, robust performance, achieving a high correlation coefficient ($r = 0.9069$, $R^2 = 0.8224$) , indicating that it effectively captures the overall trend in muscle thickness. The model's accuracy in quantifying CSA was evaluated using ground-truth annotations. A strong positive correlation was found ($r = 0.9084$, $R^2 = 0.8251$), indicating the model is highly effective at tracking CSA trends.

The model's performance on the RA muscle was moderately strong. For muscle thickness, it achieved a correlation of $r = 0.7926$ and a coefficient of determination of $R^2 = 0.6282$, indicating a good, though not exceptional, ability to capture thickness trends. For cross-sectional area (CSA), the model showed a similar moderate-to-strong correlation ($r = 0.8012$, $R^2 = 0.6419$). Therefore, the model is effective at tracking general trends for abdominal muscle but lacks the precision required for individual diagnostic or monitoring purposes.

Table 2. Results of different muscle segmentations in Dice, IoU, and HD95

Muscle Group	Dice Similarity Coefficient (DSC)	Intersection Over Union (IoU)	Hausdorff Distance (95% pixels)
Rectus Femoris (159 training images and 40 testing images)	0.826 ± 0.070	0.704 ± 0.102	14± 7
Rectus Abdominis (343 training images and 38 testing images)	0.881 ± 0.050	0.787 ± 0.080	8 ± 4.5
Biceps (174 training images and 19 testing images)	0.837 ± 0.080	0.720 ± 0.118	12 ± 6
Triceps (173 training images, and 19 testing images)	0.808 ± 0.090	0.678 ± 0.127	18 ± 9
Peroneal Muscles (6570 training images and 3465 testing images)	0.920 ± 0.040	0.852 ± 0.069	4 ± 2

For the biceps, the model demonstrated a strong correlation with ground truth for both thickness ($r = 0.8258$, $R^2 = 0.6820$) and CSA ($r = 0.8700$, $R^2 = 0.7569$), suggesting robust trend-tracking capabilities. However, due to the limited number of images, the correlation is measured with fewer images than other sites. The model exhibited its weakest performance on the triceps. For muscle thickness, the correlation was relatively poor ($r = 0.4015$, $R^2 = 0.1612$), indicating the model was largely unable to capture the relationship between predicted and true values. Performance on CSA was moderately better but still suboptimal, with a correlation of $r = 0.6840$ and $R^2 = 0.4679$.

The model developed for the PLPB muscles achieved a benchmark level of performance, demonstrating exceptional accuracy, precision, and reliability. For both muscle thickness ($r = 0.9678$, $R^2 = 0.9366$) and cross-sectional area ($r = 0.9634$, $R^2 = 0.9282$), the model yielded near-perfect correlations with the ground truth. These outperformed results can be directly attributed to the advanced training methodology and the nature of the dataset. The model was trained using the nnU-Net framework, which is more advanced than U-Net and is renowned for its ability to self-configure and robustly optimize network architecture, preprocessing, and training schemes for specific medical imaging tasks. Furthermore, the training data was generated by converting ultrasound videos into sequential image frames, yielding a substantially larger, more diverse dataset than static image collections. This large-scale dataset of 3465 images, coupled with the inherent frame-to-frame variations as a natural form of data augmentation, enabled the model to learn more generalizable and robust features. The testing set involves subject-based testing, with each subject providing the largest PLPB muscle.

3.4 Correlations between thigh, abdominals, arms, and calf with bioelectric impedance analysis (BIA)

The data from the test set is used to find the correlations with BIA and sarcopenia related parameters, mainly focusing on hand grip, 5-time sit-to-stand, and calf circumference (the major physical performance parameters); and also muscle mass, bone mass, body fat, fat free mass, visceral fat, basal metabolic rate, total body water and etc, which are major indexes from the BIA, indicating the body composition conditions.

Figure 4 shows that correlations between ultrasound-derived muscle metrics and systemic parameters varied significantly across anatomical

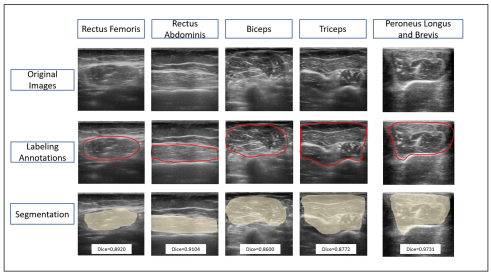

Figure 3. Examples of the results obtained from the RF, RA, Biceps, Triceps, and PLPB

sites, as indicated by p-values. For the RF, MT was moderately correlated with MM (r = 0.4224) and FFM (r = 0.4223), while its CSA showed weaker associations. Being one of the most popular site for sarcopenia measurement, many literatures has mentioned the correlation between the RF's parameter with BIA tends to be low. The moderate correlation with hand gripping results shows the association of muscles from the lower limbs to the upper limbs, i.e., the upper limb's strength, where both AWGS and EWGSOP put the hand grip into the consensus as one of the important metrics. Meanwhile, measuring the RF could demonstrate the limb's quality for sarcopenia assessment. In the light of this, ultrasound measurement of the RF muscle is used as a "window" to assess the overall muscle condition. This is because muscle, bone, water content, and basal metabolism are all closely related to the body's lean body mass. As one of the largest and most important muscles in the body, the size of the RA muscle can well reflect the overall lean body mass level. The RA demonstrated stronger relationships, with its CSA showing correlations of 0.6073 with MM and 0.6398 with basal metabolic rate (BMR). Since the around abdominal regions tend to show the visceral fat and abdominal thickness, which is also strongly related to BMR. Whether it is assessing body composition, basal metabolism or upper limb strength such as the hand grip test, the RA muscle shows to be an important muscle to describe the trunk region' muscle performance and the core stability and strength, which are fundamental to all upper and lower body movements. Therefore, the condition of your core muscles better reflects your overall function and muscular development. Upper arm muscle analysis revealed that the triceps CSA had a very strong correlation with MM (r = 0.8212) and FFM (r = 0.8243), though its association with hand grip (HG) was negligible (r = 0.1058). Similarly, the biceps MT was strongly correlated with MM (r = 0.6802) but showed little to no relationship with HG (r = 0.0608). The most robust correlations were observed for the peroneus longus, where MT was highly correlated with MM (r = 0.7537), BM (r = 0.7480), and the key sarcopenia indicator ASM (r = 0.7462).

Thigh	MT	CSA
HG	0.3937(p=0.0239)	0.5426(p=0.0501)
5XSTS	0.2399(p=0.8251)	0.0409(p=0.7957)
TBW	0.4037(p=0.0241)	0.3826(p=0.0319)
BM	0.4584(p=0.0099)	0.3824(p=0.0337)
MM	0.4224(p=0.0179)	0.3844(p=0.0327)
BMR	0.4195(p=0.0188)	0.3708(p=0.0400)
FFM	0.4223(p=0.0179)	0.3840(p=0.0328)
Calf Cir	0.1884(p=0.2557)	0.1884(p=0.3181)
FATPc	0.2835(p=0.1514)	0.2810(p=0.1245)
PM	0.0523(p=0.7789)	0.0787(p=0.6758)

BA	MT	CSA
HG	0.5905(p=0.0018)	0.5369(p=0.0014)
5XSTS	0.3043(p=0.0862)	0.2487(p=0.1558)
TBW	0.5459(p=0.0008)	0.5700(p=0.0004)
BM	0.5826(p=0.001)	0.6021(p=0.001)
MM	0.5984(p=0.0002)	0.6073(p=0.0001)
BMR	0.5349(p=0.001)	0.6398(p=0.0000)
FFM	0.5983(p=0.0002)	0.6076(p=0.0001)
Calf Cir	0.5003(p=0.0022)	0.5363(p=0.0013)
FATPc	0.4715(p=0.0049)	0.4796(p=0.0041)
PM	0.2504(p=0.1899)	0.2285(p=0.2594)
MA	0.6012(p=0.0002)	0.5984(p=0.0006)

Biceps	MT	CSA
HG	0.1213 (p=0.4282)	0.1358 (p=0.7795)
5XSTS	0.5920 (p=0.0008)	0.1143 (p=0.7539)
TBW	0.5889 (p=0.0753)	0.7336 (p=0.0180)
BM	0.5520 (p=0.0976)	0.7628 (p=0.0103)
MM	0.7969 (p=0.0072)	0.8212 (p=0.0036)
BMR	0.5401 (p=0.1875)	0.7562 (p=0.0114)
FFM	0.7772 (p=0.0085)	0.8243 (p=0.0034)
Calf Cir	0.5865 (p=0.2700)	0.4236 (p=0.2153)
VF	0.7214 (p=0.0085)	0.537 (p=0.0584)

Triceps	MT	CSA
HG	0.0608(p=0.8872)	0.1058(p=0.7713)
5XSTS	0.0424 (p=0.8644)	0.1717 (p=0.6590)
TBW	0.4749(p=0.1858)	0.4673(p=0.1702)
BM	0.5671(p=0.0873)	0.4833(p=0.1601)
MM	0.6802(p=0.0304)	0.5784(p=0.0798)
FFM	0.6788(p=0.0309)	0.5771(p=0.0805)
Calf Cir	0.1618(p=0.6543)	0.0751(p=0.8342)
VF	0.4513(p=0.1804)	0.5057 (p=0.1358)

Peroneus Longus and Brevis	MT	CSA
TBW	0.7087(p=0.0069)	0.6574(p=0.0095)
BM	0.7480(p=0.0019)	0.6936(p=0.0142)
MM	0.7537(p=0.0019)	0.7017(p=0.0104)
BMR	0.6498(p=0.0162)	0.6468(p=0.0129)
FFM	0.7354(p=0.0041)	0.7244 (p=0.0061)
Calf Cir	0.6799(p=0.0106)	0.5655(p=0.0400)
ASM	0.7462(p=0.0034)	0.6980(p=0.0208)

Figure 4. Correlations of the multi-site muscles with BIA and sarcopenia-related factors (bold for p-value < 0.05)

For the thigh (RF) muscle, the parameters with positive correlations with MT are FFM, MM, BM, TBW, HG, Calf Circumference, and 5xSTS; and FFM, MM, BM, TBW, HG, and Calf Circumference with CSA (*Figure 4*). The RF is one of the most important indicators for assessing overall muscle mass, physical function, and the risk of sarcopenia. As one of the largest and strongest muscle groups in the human body, the size of the quadriceps (which contains RF) strongly correlates with FFM, MM, and BM. This is because the quadriceps alone constitute a significant portion of these metrics. This is a key difference between the quadriceps and other muscles. It is not only significantly correlated with the 5xSTS test, which represents lower-body explosive power, such as standing up, primarily relying on thigh strength, but also with HG, a measure of upper-body strength.

For the abdominal (RA) muscle, the parameters with positive correlations to MT are MM, FFM, TBW, and Waist Circumference; and MM, FFM, TBW, Waist Circumference, VF, Fat Percentage to CSA (Table 6). The size of the RA not only reflects core muscle mass but also uniquely correlates with indicators of central adiposity. The RA correlates with MM and FFM, indicating that core muscle health is integral to the overall muscular system. RA is significantly correlated with waist circumference and VF. There are two possible reasons for this: Firstly, individuals with larger waists and more abdominal fat may require greater core strength to maintain trunk stability, and therefore may be more muscular; Secondly, the presence of abdominal fat may affect ultrasound measurement accuracy, or this correlation may reflect synchronized changes in core fat and muscle.

For the biceps muscle, the parameters with positive correlations to MT and also significant are FFM, MM, BM, TBW, BMR, and VF; and FFM, MM, BM, and TBW with CSA (*Figure 4*). Bicep's size is a good indicator of overall muscle mass and upper limb strength. As major upper limb

muscles, biceps MT and CSA are significantly correlated with FFM-free, MM, TBW, and BM, consistent with physiological expectations. This is because muscle is the primary component of FFM, contains approximately 75% water (which determines TBW), and muscle development generally correlates with skeletal strength.

For the triceps muscle, the parameters with significant positive correlations with MT and CSA are FFM, MM, and BM (*Figure 4*). Triceps size primarily reflects overall lean body mass. Unlike the biceps, triceps size was significantly correlated only with body composition indices such as FFM, MM, and BM. Triceps size did not show significant correlations with grip strength or the sit-stand test.

For the calf PLPB muscle, the parameters with positive correlations to MT and CSA are TBW, BM, MM, BMR, FFM, Muscle Index (MM/H^2), and ASM (*Figure 4*). The PLPB is an extremely reliable and sensitive measure of whole-body lean body mass and metabolic status. PLPB ultrasound parameters show very high correlation with all other parameters, including BW, BM, MM, BMR, FFM, and ASM. This indicates that changes in the size of this muscle are closely synchronized with overall muscle, bone, and metabolic status. This strong correlation likely indicates that measurements of the PLPB are highly reproducible and less susceptible to confounding factors, such as subcutaneous fat, and thus more accurately reflect true changes in the body. It can be considered an ideal "biomarker" for tracking changes in body composition.

Discussions and limitations

This study shows that the deep learning frameworks U-Net and nnU-Net have successfully developed and validated an automated segmentation method for multisite muscle assessment. This process of auto-segmentation for significant muscles in sarcopenia diagnosis shows the efficacy of ultrasound usage in sarcopenia screening. With AI-powered approaches, we could not only accurately segment and distinguish the targeted muscles like an expert physiatrist, but also demonstrate the potential for greater efficiency in sarcopenia assessment. This work directly addresses the fundamental barriers that have historically prevented ultrasound from being widely adopted for large-scale sarcopenia screening, despite its inherent advantages in assessing muscle quality and quantity.

Advantages of ultrasound aligning with deep learning methods

Historically, ultrasound has been questioned for its accuracy in assessing muscle in the evaluation of sarcopenia. Previous studies have explored deep learning for the segmentation of muscle ultrasound images. Still, they have often been limited to a single muscle group, typically the RF, or required significant manual initialization (semi-automated). Our framework represents a significant step forward by offering a fully automated, multi-site solution that covers key muscles of the upper, trunk, and lower limbs, offering a more holistic view of systemic muscle health. The core contribution of our research lies in overcoming operator dependency, subjectivity, and inefficiency. Under most clinical cases, medical ultrasound doctors tend to diagnose sarcopenia on subjective interpretation and find it hard to align the protocol. Manual drawing of muscle boundaries on ultrasound is subjective, varying between operators and even for the same operator over time. With the adoption of the deep learning framework for AI, we can provide a deterministic, objective segmentation for any given image, thereby minimizing subjectivity by eliminating inter- and intra-operator variability. By aligning this automated analysis with a standardized imaging protocol that incorporates both static and dynamic scanning, our study presents a complete, end-to-end standardized workflow. This is a critical step towards generating reproducible, comparable data needed to establish normative reference values and reliably track muscle health longitudinally.

One of the advantages of ultrasound compared with other approaches is that its efficiency is demonstrated in real time, with instant results for repetitive monitoring, and it is renowned in modern imaging technologies. Magnetic resonance imaging (MRI), though accurate and serving as the gold standard for sarcopenia evaluation, is time-consuming and expensive. The bulky machine could be inconvenient for repetitive measurements (Codari et al., 2020; Sergi et al., 2016). The same for the dual energy x-ray absorptiometry (DXA), though recommended by both AWGS and EWGSOP, DXA is usually expensive and has risks of radiation, though smaller compared to CT, is also not an ideal method for long-term monitoring, especially for community, elderly people's house and in-home settings since the most prohibitive barrier to using ultrasound in community settings is the time-intensive nature of manual analysis (Albano et al., 2020; Guglielmi et al., 2016; Zhang et al.).

Multi-site scanning and segmentations

Our study's primary contribution is the development of a highly accurate, automated framework for segmenting multiple, functionally diverse muscles. The clinical significance of this multi-site approach is best understood by examining the distinct roles of each tar-

geted muscle group in the context of sarcopenia and geriatric health. The RF, as a primary component of the quadriceps, is fundamental for lower-limb power and mobility(Morley et al., 2011). It is essential for critical activities of daily living such as walking, rising from a chair, and climbing stairs(Lauretani et al., 2003). Its relatively superficial location and clear fascial boundaries have made it a common target for ultrasound-based sarcopenia research. Current models enable clinicians and healthcare providers to rapidly assess a key muscle related to mobility and lower-body strength, facilitating early detection of mobility impairments. Upper-limb sarcopenia directly compromises an individual's capacity to perform essential self-care tasks, including lifting objects, dressing, and maintaining personal hygiene (Dohertya, 2010). The biceps and triceps are not only crucial for these functions but are also primary targets for resistance training interventions aimed at combating muscle decline (Feng et al., 2024; Harris-Love et al., 2018). The ability of our framework to precisely track morphological changes in these muscles offers a powerful tool for monitoring the efficacy of exercise regimens or nutritional therapies, enabling personalized adjustments to treatment plans and objectively quantifying patient progress. In this study, the triceps have lower and fewer correlations, possibly because the triceps do not play a more direct role in these specific tests than other muscle groups (such as the forearms or legs). The RA is a key element of the core muscular region, which is vital for trunk stability, posture, and fundamental transfer movements (Tagliafico et al., 2022). A decline in core strength is often implicated when older adults struggle to move from a supine to a sitting position (e.g., getting out of bed), a critical marker of functional decline and dependency(Kato & Hatanaka, 2020). By offering an objective measure of this key trunk muscle, our AI-driven applications can help identify core muscle deficits and track the outcomes of targeted core strengthening programs, which are crucial for maintaining functional independence and preventing falls (Sasaki et al., 2018).

The most novel aspect of our framework is the successful and highly accurate segmentation of the anatomically complex peroneal muscle group of PLPB. While larger leg muscles like the gastrocnemius are typically assessed for power, the peroneal muscles play a critical, yet often overlooked, role in dynamic ankle stability and postural control(Fan et al., 2022). Their proper function is essential for a person's ability to maintain balance on uneven surfaces and to react to unexpected postural challenges(Perez-Sousa et al., 2019). Dysfunction in these muscles is directly linked to an increased risk of trips and falls (Bavdek et al., 2018; Santilli et al., 2005). The exceptional performance of our nnU-Net model on this task not only highlights its technical robustness but, more importantly, opens a new avenue for a more comprehensive, function-oriented assessment of fall risk that goes beyond simple muscle mass measurement.

By addressing these practical bottlenecks, our study demonstrates that integrating a deep learning framework from the upper limbs to the trunk and from the trunk to the lower limbs enables the unlocking of the actual clinical potential of ultrasound, thereby paving the way for its widespread adoption in geriatric care. The most immediate implication is the enablement of large-scale, community-based sarcopenia screening. Mobile health equipment, such as local clinics equipped with portable ultrasound devices and our software, can efficiently screen vast elderly populations, identifying at-risk individuals long before the onset of overt functional decline, including falls or immobility. This proactive approach aligns perfectly with the goals of preventative geriatrics.

Clinical significance for the multi-site scanning of ultrasound incorporating deep learning methods

The core clinical significance of this study lies in its validation of an automated ultrasound muscle assessment methodology based on the U-Net and nnU-Net models, with Dice accuracy exceeding 0.8 for the major muscle assessment. The results demonstrate that muscle metrics, such as MT and CSA, derived from deep learning-driven automatic segmentation of ultrasound images, exhibit a noticeable agreement with BIA-generated parameters, an established gold standard for body composition. The strength of this correlation is comparable to, and exceeds, that achieved by traditional manual tracing methods, conferring multiple layers of clinical value. We have shown that the major muscles of the upper and lower limbs, from the trunk to the thigh, encompassing the overall assessment of the major muscle groups of humans, are capable of providing a complete understanding of sarcopenia.

Primarily, this approach enables the automation and standardization of the muscle assessment workflow. Conventional manual measurement is not only time-consuming but also highly dependent on the operator's experience, leading to significant inter- and intra-operator variability. The introduction of AI models fundamentally reduces this source of human error by providing an objective, consistent segmentation standard, thereby ensuring the reproducibility and reliabil-

ity of results. Secondly, workflow automation dramatically enhances clinical efficiency, enabling rapid and accurate muscle assessment in busy clinical settings and large-scale epidemiological screenings.

More importantly, the U-Net and nnU-Net deep learning models help improve the accessibility of healthcare services. By embedding expert-level knowledge into an algorithm, it lowers the skill threshold for ultrasound operation, enabling high-quality screening and monitoring for sarcopenia and other muscle-related conditions in primary care or resource-limited settings using portable ultrasound devices coupled with AI software. In summary, this research does not merely validate the performance of an AI model; it unveils a powerful new clinical tool that can transform ultrasound muscle assessment into a faster, more precise, and objective standardized process, thereby offering a novel "digital biomarker" for the management of muscle health and the diagnosis of related diseases.

Furthermore, the high efficiency and reproducibility of our tool make it ideal for longitudinal and repetitive monitoring. Clinicians can now feasibly perform frequent, low-cost ultrasound scans to quantitatively track a patient's response to nutritional or exercise interventions. Additionally, this tool supports a shift from reactive treatment to proactive prevention in geriatric healthcare. By enabling frequent, non-invasive monitoring of muscle health, clinicians can track the effectiveness of interventions—such as nutritional support or resistance training—in real-time. This aligns with the aim of gerontechnology to enhance the quality of life for older adults through technological innovation. It transforms sarcopenia assessment from a rare, hospital-based, diagnostic event into a routine health metric, much like blood pressure monitoring, ensuring that older adults receive timely care to maintain their physical independence.

Limitations and future usage of deep learning in ultrasound for sarcopenia assessment

This study has several limitations. First, we only include the major muscle groups for sarcopenia-related physical activity for the elderly. However, in the future, we are eager to include more important muscle groups such as the gastrocnemius. Meanwhile, we could include gender differences in the study. Future work should involve testing the framework on external datasets from different institutions and ultrasound machines with different probes of various models. Second, while our AI demonstrates excellent agreement with an expert, we did not compare the derived ultrasound measurements with a "gold standard" for muscle mass, such as MRI or DXA. Such a comparison would be a valuable next step to establish the criterion validity of our AI-derived biomarkers. Finally, a future longitudinal study is required to determine its predictive power for clinical outcomes such as falls, frailty, and mortality, and additional muscle studies relevant to geriatric health are also needed for future ultrasound-based screening.

Conclusions

In conclusion, our study has developed and validated an AI-powered framework that transforms musculoskeletal ultrasound into a rapid, objective, and standardized tool for overall sarcopenia assessment, and the predicted results proved to have good agreement with BIA-generated body composition parameters. By systematically dismantling the key barriers to its clinical implementation, this work paves the way for the widespread adoption of ultrasound in community screening, personalized monitoring, and preventative geriatric intervention. This technological solution exemplifies the core mission of gerontechnology: to develop practical, scalable tools that enhance the health and well-being of aging populations.

Funding information

This research (As well as for the APC) was funded by the Research Institute for Smart Ageing (RISA) of The Hong Kong Polytechnic University (1-CDK0). The project was also supported by the Hong Kong University Grants Council (ref. no. 15605822).

Ethical approval

Ethical approval for this study was obtained from the Human Subject Ethics Sub-committee (HSESC) of the Hong Kong Polytechnic University, HSEARS 20250528002 and HSEARS 20250610004

References

Abe, T., Nakatani, M., & Loenneke, J. P. (2018). Relationship between ultrasound muscle thickness and MRI-measured muscle cross-sectional area in the forearm: a pilot study. Clinical Physiology and Functional Imaging, 38(4), 652-655.

Albano, D., Messina, C., Vitale, J., & Sconfienza, L. M. (2020). Imaging of sarcopenia: old evidence and new insights. European radiology, 30(4), 2199-2208.

Bavdek, R., Zdolšek, A., Strojnik, V., & Dolenec, A. (2018). Peroneal muscle activity during different types of walking. Journal of Foot and Ankle Research, 11(1), 50.

Behboodi, B., Obrand, J., Afilalo, J., & Rivaz, H. (2024). Deepsarc-us: A deep learning framework for assessing sarcopenia using ultrasound images. Applied Sciences, 14(15), 6726.

Cao, L., & Morley, J. E. (2016). Sarcopenia is recognized as an independent condition by an inter-

national classification of disease, tenth revision, clinical modification (ICD-10-CM) code. Journal of the American Medical Directors Association, 17(8), 675-677.

Chang, K.-V., Wu, W.-T., Huang, K.-C., Jan, W. H., & Han, D.-S. (2018). Limb muscle quality and quantity in elderly adults with dynapenia but not sarcopenia: an ultrasound imaging study. Experimental Gerontology, 108, 54-61.

Chen, L.-K., Liu, L.-K., Woo, J., Assantachai, P., Auyeung, T.-W., Bahyah, K. S., Chou, M.-Y., Chen, L.-Y., Hsu, P.-S., & Krairit, O. (2014). Sarcopenia in Asia: consensus report of the Asian Working Group for Sarcopenia. Journal of the American Medical Directors Association, 15(2), 95-101.

Chen, L.-K., Woo, J., Assantachai, P., Auyeung, T.-W., Chou, M.-Y., Iijima, K., Jang, H. C., Kang, L., Kim, M., & Kim, S. (2020). Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. Journal of the American Medical Directors Association, 21(3), 300-307. e302.

Chen, Y. L., Liu, P. T., Chiang, H. K., Lee, S. H., Lo, Y. L., Yang, Y. C., & Chiou, H. J. (2022). Ultrasound Measurement of Rectus Femoris Muscle Parameters for Discriminating Sarcopenia in Community-Dwelling Adults. Journal of Ultrasound in Medicine, 41(9), 2269-2277.

Chen, Z.-T., Li, X.-L., Jin, F.-S., Shi, Y.-L., Zhang, L., Yin, H.-H., Zhu, Y.-L., Tang, X.-Y., Lin, X.-Y., & Lu, B.-L. (2025). Diagnosis of sarcopenia using convolutional neural network models based on muscle ultrasound images: prospective multicenter study. Journal of Medical Internet Research, 27, e70545.

Cho, M.-R., Lee, S., & Song, S.-K. (2022). A review of sarcopenia pathophysiology, diagnosis, treatment and future direction. Journal of Korean medical science, 37(18).

Codari, M., Zanardo, M., di Sabato, M. E., Nocerino, E., Messina, C., Sconfienza, L. M., & Sardanelli, F. (2020). MRI-derived biomarkers related to sarcopenia: A systematic review. Journal of Magnetic Resonance Imaging, 51(4), 1117-1127.

Cruz-Jentoft, A. J., Baeyens, J. P., Bauer, J. M., Boirie, Y., Cederholm, T., Landi, F., Martin, F. C., Michel, J.-P., Rolland, Y., & Schneider, S. M. (2010). Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age and ageing, 39(4), 412-423.

Cruz-Jentoft, A. J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., & Sayer, A. A. (2019a). Sarcopenia: revised European consensus on definition and diagnosis. Age and ageing, 48(1), 16-31.

Cruz-Jentoft, A. J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., & Sayer, A. A. (2019b). Writing Group for the European Working Group on Sarcopenia in Older People 2 (EWGSOP2), and the Extended Group for EWGSOP2. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing, 48(1), 16-31.

Dohertya, M. J. B. T. J. (2010). Sarcopenia: prevalence, mechanisms, and functional consequences. Body composition and aging, 37, 94.

Fan, Y., Zhang, B., Huang, G., Zhang, G., Ding, Z., Li, Z., Sinclair, J., & Fan, Y. (2022). Sarcopenia: Body composition and gait analysis. Frontiers in aging neuroscience, 14, 909551.

Feng, T., Zhao, C., Dong, J., Xue, Z., Cai, F., Li, X., Hu, Z., & Xue, X. (2024). The effect of unaffected side resistance training on upper limb function reconstruction and prevention of sarcopenia in stroke patients: a randomized controlled trial. Scientific Reports, 14(1), 25330.

Fuchs, C. J., Kuipers, R., Rombouts, J. A., Brouwers, K., Schrauwen-Hinderling, V. B., Wildberger, J. E., Verdijk, L. B., & van Loon, L. J. (2023). Thigh muscles are more susceptible to age-related muscle loss when compared to lower leg and pelvic muscles. Experimental Gerontology, 175, 112159.

Guglielmi, G., Ponti, F., Agostini, M., Amadori, M., Battista, G., & Bazzocchi, A. (2016). The role of DXA in sarcopenia. Aging Clinical and Experimental Research, 28(6), 1047-1060.

Guo, J.-Y., Zheng, Y.-P., Kenney, L. P., Bowen, A., Howard, D., & Canderle, J. J. (2011). A comparative evaluation of sonomyography, electromyography, force, and wrist angle in a discrete tracking task. Ultrasound in medicine & biology, 37(6), 884-891.

Harris-Love, M. O., Benson, K., Leasure, E., Adams, B., & McIntosh, V. (2018). The influence of upper and lower extremity strength on performance-based sarcopenia assessment tests. Journal of functional morphology and kinesiology, 3(4), 53.

Hida, T., Ando, K., Kobayashi, K., Ito, K., Tsushima, M., Kobayakawa, T., Morozumi, M., Tanaka, S., Machino, M., & Ota, K. (2018). Ultrasound measurement of thigh muscle thickness for assessment of sarcopenia. Nagoya journal of medical science, 80(4), 519.

Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., & Wirkert, S. (2018). nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486.

Kato, K., & Hatanaka, Y. (2020). The influence of trunk muscle strength on walking velocity in elderly people with sarcopenia. Journal of Physical Therapy Science, 32(2), 166-172.

Kuyumcu, M. E., Halil, M., Kara, Ö., Çuni, B., Çağlayan, G., Güven, S., Yeşil, Y., Arık, G., Yavuz, B. B., & Cankurtaran, M. (2016). Ultrasonographic evaluation of the calf muscle mass and architecture in elderly patients with and without sarcopenia. Archives of gerontology and geriatrics, 65, 218-224.

Larsson, L., Degens, H., Li, M., Salviati, L., Lee, Y. I., Thompson, W., Kirkland, J. L., & Sandri, M. (2019). Sarcopenia: aging-related loss of muscle mass and function. Physiological reviews, 99(1), 427-511.

Lauretani, F., Russo, C. R., Bandinelli, S., Bartali, B., Cavazzini, C., Di Iorio, A., Corsi, A. M., Rantanen, T., Guralnik, J. M., & Ferrucci, L. (2003). Age-

associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. Journal of applied physiology, 95(5), 1851-1860.

Li, S., Li, H., Hu, Y., Zhu, S., Xu, Z., Zhang, Q., Yang, Y., Wang, Z., & Xu, J. (2020). Ultrasound for measuring the cross-sectional area of biceps brachii muscle in sarcopenia. International Journal of Medical Sciences, 17(18), 2947.

Morley, J. E., Abbatecola, A. M., Argiles, J. M., Baracos, V., Bauer, J., Bhasin, S., Cederholm, T., Coats, A. J. S., Cummings, S. R., & Evans, W. J. (2011). Sarcopenia with limited mobility: an international consensus. Journal of the American Medical Directors Association, 12(6), 403-409.

Nies, I., Ackermans, L., Poeze, M., Blokhuis, T., & Ten Bosch, J. A. (2022). The diagnostic value of ultrasound of the rectus femoris for the diagnosis of sarcopenia in adults: a systematic review. Injury, 53, S23-S29.

Perez-Sousa, M. A., Venegas-Sanabria, L. C., Chavarro-Carvajal, D. A., Cano-Gutierrez, C. A., Izquierdo, M., Correa-Bautista, J. E., & Ramírez-Vélez, R. (2019). Gait speed as a mediator of the effect of sarcopenia on dependency in activities of daily living. Journal of cachexia, sarcopenia and muscle, 10(5), 1009-1015.

Rolland, Y., Lauwers-Cances, V., Cournot, M., Nourhashémi, F., Reynish, W., Rivière, D., Vellas, B., & Grandjean, H. (2003). Sarcopenia, calf circumference, and physical function of elderly women: a cross-sectional study. Journal of the American Geriatrics Society, 51(8), 1120-1124.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention,

Santilli, V., Frascarelli, M. A., Paoloni, M., Frascarelli, F., Camerota, F., De Natale, L., & De Santis, F. (2005). Peroneus longus muscle activation pattern during gait cycle in athletes affected by functional ankle instability: a surface electromyographic study. The American journal of sports medicine, 33(8), 1183-1187.

Sasaki, E., Sasaki, S., Chiba, D., Yamamoto, Y., Nawata, A., Tsuda, E., Nakaji, S., & Ishibashi, Y. (2018). Age-related reduction of trunk muscle torque and prevalence of trunk sarcopenia in community-dwelling elderly: Validity of a portable trunk muscle torque measurement instrument and its application to a large sample cohort study. PloS one, 13(2), e0192687.

Sergi, G., Trevisan, C., Veronese, N., Lucato, P., & Manzato, E. (2016). Imaging of sarcopenia. European journal of radiology, 85(8), 1519-1524.

Tagliafico, A. S., Bignotti, B., Torri, L., & Rossi, F. (2022). Sarcopenia: how to measure, when and why. La radiologia medica, 127(3), 228-237.

Yi, J., Shin, Y., Hahn, S., & Lee, Y. H. (2022). Deep learning based sarcopenia prediction from shear-wave ultrasonographic elastography and gray scale ultrasonography of rectus femoris muscle. Scientific Reports, 12(1), 3596.

Zhang, D., Kang, H., Sun, Y., Liu, J. Y. W., Lee, K.-S., Song, Z., Khaw, J. V., Yeung, J., Peng, T., & Lam, S.-k. (2024). Rectus Femoris Muscle Segmentation on Ultrasound Images of Older Adults Using Automatic Segment Anything Model, nnU-Net and U-Net—A Prospective Study of Hong Kong Community Cohort. Bioengineering, 11(12), 1291.

Zhang, D., Lam, S. K., & Zheng, Y. A Comprehensive Review of the Correlations of Measurement Parameters among Modern Technologies for Sarcopenia Assessment. Aging and disease.

Zheng, Y. P., Lee, T. T. Y., Lai, K. K. L., Yip, B. H. K., Zhou, G. Q., Jiang, W. W., ... & Lam, T. P. (2016). A reliability and validity study for Scolioscan: a radiation-free scoliosis assessment system using 3D ultrasound imaging. Scoliosis and spinal disorders, 11(1), 13.