Appendicular Lean Mass Estimation from Dual-View DXA Scans Using Hierarchical Fusion

Marion Mundt[1,2,3] ✉, Afsah Saleem[1,2], Abadi K Gebre[1], Arooba Maqsood[1,2], John T Schousboe[4,5], William D Leslie[6], Nicholas C Harvey[7,8], Syed Zulqarnain Gilani[1,2,9], Kun Hu[2], David Suter[1,2], Joshua R Lewis[1,2,10], and Marc Sim[1,2,10]

[1] Nutrition & Health Innovation Research Institute, School of Medical and Health Sciences, Edith Cowan University, Perth, WA, Australia
m.mundt@ecu.edu.au
[2] Centre for Artificial Intelligence and Machine Learning, School of Science, Edith Cowan University, Perth, WA, Australia
[3] School of Human Sciences, The University of Western Australia, Perth, WA, Australia
[4] Park Nicollet Clinic and HealthPartners Institute, Minneapolis, USA
[5] Division of Health Policy and Management, University of Minnesota, Minneapolis, USA
[6] Department of Internal Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
[7] MRC Lifecourse Epidemiology Centre, University of Southampton, Southampton, UK
[8] NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK
[9] Department of Computer Science and Software Engineering, The University of Western Australia, Perth, WA, Australia
[10] Medical School, The University of Western Australia, Perth, WA, Australia

Abstract. Low appendicular lean mass (ALM) is a biomarker that is associated with an increased risk of weakness, falls, functional decline, frailty, and adverse health outcomes, especially when combined with impaired strength or physical performance. ALM is usually determined using whole-body dual-energy X-ray absorptiometry (DXA) scans or by means of bioelectrical impedance analysis. However, both methods are uncommonly performed in routine clinical care. We therefore develop a deep learning model that employs hierarchical fusion mechanisms to use clinical data and DXA scans of hip and spine to estimate ALM. These scans are already collected during routine osteoporosis screening in millions of people worldwide. The proposed model uses two independent image encoders and a clinical data encoder to extract relevant features from each modality. These features are fused hierarchically using bidirectional feature-wise linear modulation. The combined framework outperforms single modalities by up to 47% in a cohort of more than 30,000 individuals. The clinical relevance is established by showing cross-sectional associations between predicted ALM and hand grip strength in more than 50,000 individuals. These results are the first step towards a scalable tool for ALM estimation.

Keywords: Falls and fracture · Sarcopenia · Osteoporosis Screening · Hierarchical Fusion · Multimodal Data.

1 Introduction

Low appendicular lean muscle mass (ALM) is a biomarker often observed in older adults, that is associated with an increased risk of weakness, functional decline, frailty, and adverse health outcomes, including reduced quality of life, a higher risk of falls and fractures, and a higher risk of mortality especially when combined with impaired strength or physical performance [4, 2]. Considering the increased global life expectancy from 64.9 to 73.3 years over the past two decades and the number of people aged 60 years and older growing from 1.1 billion in 2023 to a projected 1.4 billion by 2030, strategies to minimise age related functional decline and to support healthy ageing are needed [21]. However, despite its clinical relevance, ALM is rarely available in routine care, limiting large-scale screening for loss of ALM and longitudinal monitoring in ageing populations.

ALM can be measured using whole-body dual-energy X-ray absorptiometry (DXA), which is considered the reference standard, or by bioelectrical impedance analysis (BIA), which is a portable and cost-efficient alternative [3]. However, BIA relies on proprietary equations, which limits between-devices comparability and often results in an overestimation of ALM [1, 6]. Despite their reliability, whole-body DXA and BIA are not part of routine clinical practice and often excluded from healthcare benefits.

To make ALM monitoring more accessible, several attempts have been made to estimate ALM without the use of whole-body DXA scans or BIA. Conventional machine learning algorithms were trained on clinical and anthropometric data, that is not being collected in routine clinical practice using small datasets with limited generalisability, e.g. [12, 7, 18]. A first application of using imaging for ALM estimation was made using a machine learning model based on VGG16 pretrained on ImageNet on a small sample of chest X-rays and anthropometric data [15]. Further, routinely captured DXA images of the lateral spine have been used for the estimation of visceral fat using MobileNet v2 as image encoder [11]. These previous results show that cost-efficient imaging modalities like X-ray or DXA include information on different tissue and allow for the estimation of tissue mass that is not captured within the image boundaries, although they face the challenge of low signal-to-noise ratio, low contrast, and image artefacts [11].

In this study, we investigate the prediction of ALM from hip, anterior-posterior (AP) spine DXA scans, and routine clinical parameters: sex, age, and weight. Parameters derived from these images in combination with clinical inputs have shown high accuracy in ALM estimation using multiple linear regression in small cohorts [14, 20]. However, hip%fat and spine%fat used in these predictive equations are not standard DXA outputs and therefore not available in routine practice. To directly integrate images, we developed a machine learning model that consists of three branches, two vision encoders and a tabular encoder with two multi-model fusion mechanisms to estimate ALM. We further

validated the approach by analysing the association of machine learning derived ALM (ML_ALM) and ground-truth ALM with hand grip strength. The main contributions of this paper are:

- This is the first study to directly use DXA images that are routinely collected during osteoporosis screening in a high-risk population for functional decline and provide information linked to lean mass and muscle strength. This enables information on bone and muscle to be provided concurrently.
- We develop a hierarchical fusion approach to combine complementary information from hip and AP spine DXA scans with clinical data to estimate ALM in the UK Biobank Imaging Study [10], comprising of more than 30,000 DXA scans.
- We validate the predicted measures against hand grip strength in a cohort of more than 50,000 individuals.

2 Methodology

Given dual-view DXA images and associated clinical variables, the proposed framework learns a compact multimodal representation for downstream prediction. We aim to learn an optimal joint embedding by integrating complementary information from multiple imaging views and clinical features through hierarchical fusion mechanisms (Figure 1).

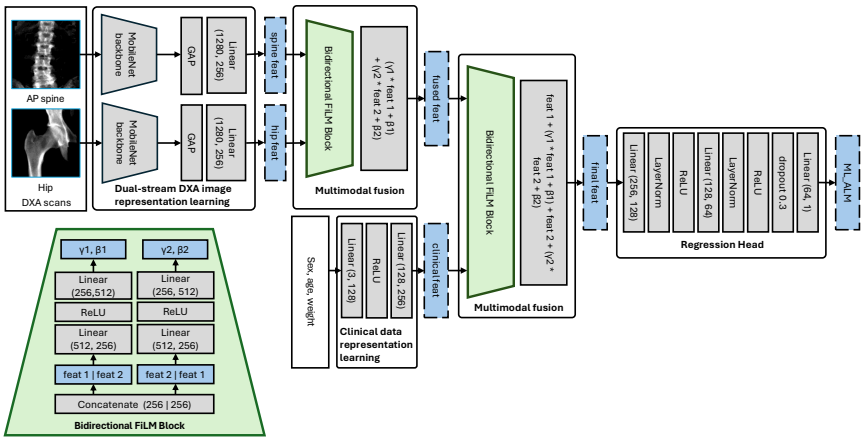


Fig. 1: The framework consists of four modules: i) dual-stream DXA image representation learning, ii) clinical data representation learning, iii) multimodal fusion across imaging and clinical data, and iv) regression head for outcome prediction.

Dual-stream DXA Image Representation Learning: Given a hip DXA image X^{hip} and an AP spine DXA image X^{spine}, generic feature maps are extracted using CNN backbone encoders such that $\psi^{\text{i}} = f_{\theta^{\text{i}}}(X^{\text{i}}), \quad i \in \{\text{hip}, \text{spine}\}$, where $\psi \in \mathbb{R}^{h \times w \times c}$, and h, w, and c denote the spatial resolution and channel dimensionality of the feature maps. We employ light-weight CNNs as image encoders,

as DXA images exhibit strong spatial structure and limited texture variability, which align well with CNN's inductive biases [22]. Compared to transformer-based models, CNNs provide more stable optimisation on moderate-sized medical datasets, lower computational cost, and efficient inference in multi-view settings [19, 11]. Separate encoders with identical architectures are employed to preserve image-specific anatomical characteristics of the hip and AP spine views. The final classification layers of a MobileNetV2 model [16] pretrained on ImageNet [5] are removed, and Global Average Pooling (GAP) is applied to the convolutional feature maps to obtain fixed-length vector representations [11], $\mathbf{f}_{\mathrm{i}} = \mathrm{GAP}(\psi^{\mathrm{i}}), \quad i \in \{\mathrm{hip}, \mathrm{spine}\}$, where $\mathbf{f}_{\mathrm{i}} \in \mathbb{R}^{256}$ denote the final hip and AP spine image embeddings.

Clinical Feature Representation Learning: In parallel, structured clinical variables $Z \in \mathbb{R}^{d_z}$ are encoded into a latent representation $\phi = g_\theta(Z)$, where d_z denotes the dimensionality of the clinical feature vector Z. Anthropometric features (sex, age, weight) are considered as inputs to the tabular encoder. A multilayer perceptron (MLP) is employed for clinical feature encoding due to its strong performance and training stability on clinical datasets, computational efficiency, and straightforward integration with image-derived representations. Although transformer-based tabular models can explicitly capture feature interactions [11], no performance gain was observed in our setting. The clinical feature vector $Z \in \mathbb{R}^{d_z}$ is encoded using a two-layer multilayer perceptron (MLP) with ReLU activations. Specifically, Z is first projected into a 128-dimensional hidden representation, followed by a ReLU activation, and subsequently mapped to a 256-dimensional clinical embedding $\mathbf{f}_{\mathrm{clinical}} \in \mathbb{R}^{256}$.

Hierarchical Multimodal Fusion: We adopt a hierarchical fusion strategy in which feature embedding $\mathbf{f}_{\mathrm{hip}}$ and $\mathbf{f}_{\mathrm{spine}}$ are first combined together using bidirectional feature-wise linear modulation (BiFiLM) fusion [13] to create $\mathbf{f}_{\mathrm{img}}$ as the fused image representation prior to integration with clinical features $\mathbf{f}_{\mathrm{clinical}}$. As both images share the same modality and spatial structure, early feature-level fusion facilitates learning joint anatomical representations. Clinical variables, which provide non-spatial global context, are fused using residual BiFiLM at a later stage to condition the learned visual features, improving optimisation stability in multimodal learning [9]. FiLM fusion performs feature-wise affine modulation, providing a parameter-efficient and computationally lightweight alternative to attention-based fusion, while often exhibiting stable optimisation behaviour in multimodal settings. [8, 17]. FiLM [13] performs affine conditioning of one feature representation using parameters predicted from the joint feature space is given as follows where $\boldsymbol{\gamma}, \boldsymbol{\beta} \in \mathbb{R}^d$ denote feature-wise scaling and bias terms, respectively:

$$\boldsymbol{\gamma}, \boldsymbol{\beta} = \mathrm{MLP}([\mathbf{f}_1 \,\|\, \mathbf{f}_2]),$$
$$\mathbf{z} = \boldsymbol{\gamma} \odot \mathbf{f}_1 + \boldsymbol{\beta},$$

To allow mutual conditioning between feature representations, we extend FiLM

to a bidirectional formulation in which each modality modulates the other:

$$\begin{aligned}\boldsymbol{\gamma}_1, \boldsymbol{\beta}_1 &= \mathrm{MLP}([\mathbf{f}_1 \,\|\, \mathbf{f}_2]),\\ \boldsymbol{\gamma}_2, \boldsymbol{\beta}_2 &= \mathrm{MLP}([\mathbf{f}_2 \,\|\, \mathbf{f}_1]),\\ \mathbf{z} &= (\boldsymbol{\gamma}_1 \odot \mathbf{f}_1 + \boldsymbol{\beta}_1) + (\boldsymbol{\gamma}_2 \odot \mathbf{f}_2 + \boldsymbol{\beta}_2).\end{aligned}$$

Finally, we introduce residual BiFiLM to preserve the original feature representations while enabling bidirectional modulation and improved gradient flow:

$$\mathbf{z} = \mathbf{f}_1 + \boldsymbol{\gamma}_1 \odot \mathbf{f}_1 + \boldsymbol{\beta}_1 + \mathbf{f}_2 + \boldsymbol{\gamma}_2 \odot \mathbf{f}_2 + \boldsymbol{\beta}_2.$$

Regression Head: We use a simple fully connected multilayer perceptron (MLP) regression head with two hidden layers of 128 and 64 units, each followed by ReLU activation and dropout (p=0.3). The input feature dimension is 256 and the output is a single scalar prediction for ALM.

Percentile-weighted Regression Loss: To address the imbalance and clinical relevance of clinically low values, we use a sex-stratified, percentile-weighted mean squared error (MSE) loss. Specifically, the lowest percentile within their sex distribution is assigned a higher loss weight to encourage accurate prediction of clinically low measurements while preserving overall regression performance. The loss is defined as:

$$\mathcal{L} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} w_{i,t}\big(y_{i,t} - \hat{y}_{i,t}\big)^2, \quad w_{i,t} = \begin{cases} w_{\text{high}}, & y_{i,t} < Q_p^{\text{sex}(i),t},\\ 1, & \text{otherwise,}\end{cases}$$

where N denotes the number of samples, T the number of regression targets, $y_{i,t}$ and $\hat{y}_{i,t}$ the ground-truth and predicted values for sample i and target t, respectively, and $Q_p^{\text{sex}(i),t}$ the p-th percentile of target t computed within the sex group of sample i. The parameter w_{high} controls the increased penalty assigned to low-percentile samples. In our cohort, approximately 10% of both men and women exhibit low ALM. To improve prediction accuracy for these individuals and those near the clinical threshold, we set $p = 25$ and $w_{\text{high}} = 5$, thereby emphasising samples below the sex-specific 25^{th} percentile during training.

3 Experiments

Dataset: The model was trained on low-resolution, archive DXA scans from the UK Biobank [10] (ethical approval ID 93712) captured at two instances (2014+ and 2019+) using Lunar iDXA scanners (General Electric Healthcare, WI, US). Images of the hip (proximal femur) and anterior-posterior (AP) spine were extracted along with age, weight and sex. ALM was calculated as the sum of the lean mass of the arms and legs derived from whole-body DXA images. To ensure the quality of whole-body DXA images and ALM values, any image with any body part falling outside the image parameters and their corresponding

ALM values were removed from the sample. At the time the images were provided (November 2023), a total of 33,938 samples with all the input data were available.

Processing: Hip and AP spine DXA images exhibit substantial variability in aspect ratio (mean ± SD: hip 0.93 ± 0.12, spine 1.03 ± 0.07), spatial resolution (hip: 218×280 to 910×716 pixels; spine: 275×276 to 1036×720 pixels), and backgrounds. To standardise the inputs, we (i) fill the background regions with black, (ii) crop images to the region of interest by removing rows and columns containing more than 80% black pixels, and (iii) resize all images to 224×224 pixels with padding applied to preserve the original aspect ratio, (iv) augment training images with random rotations ($\pm 10°$), translations ($\pm 10\%$), and scaling (0.6–1.1) to account for acquisition variability in DXA scans.

Model Training: The full dataset (n = 33,938) is split into 80% training/validation and a 20% test set using grouped five-fold cross-validation. Grouping is performed at the subject level to ensure that all longitudinal scans from the same individual are assigned to a single fold and each sample appears in the test set exactly once. Prior to training, clinical inputs and ALM targets are standardised to zero mean and unit variance using statistics computed from the training split of each fold. All images are normalised to the range $[-1, 1]$ using a mean and standard deviation of 0.5 and replicated across three channels to match the input requirements of the image encoder. The optimal architecture for each branch (hip image, AP spine image, and clinical data) is selected independently using one fold of the training data. Similarly, the image–image and fused–image–clinical fusion strategies are determined through this process. Training hyperparameters, including batch size and learning rate, are selected via grid search, with validation loss used to guide dynamic learning rate adjustment. All models are trained using the Adam optimiser for a maximum of 50 epochs with a batch size of 32. The initial learning rate is set to 1×10^{-4} and reduced by a factor of 0.5 if the validation loss plateaus for more than five epochs. Early stopping is applied if no improvement in validation loss is observed for 10 consecutive epochs to mitigate overfitting.

Following cross-validation, model was trained on the full dataset to allow for the prediction of ALM in individuals with complete input but without ground-truth ALM data. The model was trained for a maximum of 30 epochs without changing the learning rate. These settings represented those observed during cross-validation and allowed for training a model on the maximum data available.

Model Evaluation: Root-mean-squared error (RMSE), mean absolute percentage error (MAPE), intra-class correlation coefficient ICC(1,1), and coefficient of determination (R^2) were calculated to evaluate the agreement between predicted ML_ALM and ground-truth ALM. To avoid aggregation bias, results for men and women were presented separately. All analyses were performed using Python (v3.12). The machine learning model was developed using the pytorch package and trained on an NVIDIA GeForce RTX 3060. The statistical analysis was undertaken using the scipy and statsmodels packages.

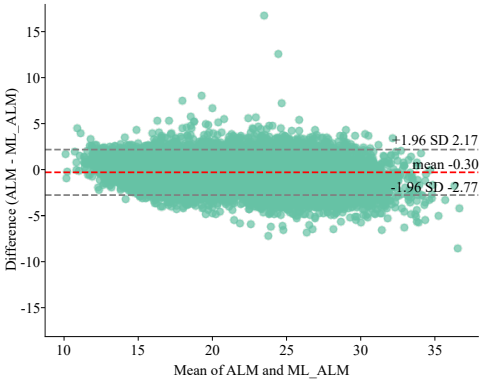
4 Results and Discussion

The final model was developed by finding the best architecture for each modality separately and identifying the best performing fusion mechanism. MobileNetV2 was used as image encoder for both hip and spine DXA images, tabular data was encoded using an MLP network, images were fused using BiFiLM, and the fused images were combined with the tabular data via residual BiFiLM, as these were the best performing fusing mechanisms compared with concatenation, fusion gate, and simple FiLM fusion. The results per modality are given in Table 1.

Table 1: Overall model performance measured by MAPE and R^2 by data modality in a 5-fold cross-validation

Data modality	MAPE [%]	R^2
Hip image only	8.19	0.770
Spine image only	9.95	0.697
Tabular data only	6.69	0.857
Hip and spine images (BiFiLM)	7.18	0.819
Hip and spine images (BiFiLM) and tabular data (residual BiFiLM)	4.92	0.924

In the 5-fold cross-validation the MAPE between ground-truth and predicted values was ∼5% for men and women. The Bland-Altman plot shows only small bias and strong agreement (Figure 2), indicating that routinely collected DXA images and clinical data incorporate complementary information that support the estimation of ALM. The proposed bidirectional fusion mechanisms for images and clinical values allow to exploit this information through mutual conditioning between modalities.



	Male	Female
N	15996	17942
RMSE [kg]	1.460	1.129
MAPE [%]	4.62	5.19
ICC(1,1)	0.875	0.860
R^2	0.774	0.744
Low ALMI (ALM/height^2)		
Accuracy		0.901
Specificity		0.938
Sensitivity		0.632

Fig. 2: Bland-Altman plot showing the agreement between ground-truth ALM and predicted ML_ALM (left). RMSE, MAPE, ICC(1,1), and R^2 (top right). Classification results for low appendicular lean mass index according to EWGSOP2 (male <7.5kg/m^2, female <5.5kg/m^2) (bottom right)

We have not undertaken a direct comparison with other existing machine learning models because (i) none have used DXA images as direct inputs to their model, (ii) these models use non-standard input variables that are not available in our dataset, and (iii) the published models included less than 2,500 participants, while our dataset includes more than 30,000 samples. Such smaller datasets may have less variability or allow the model to memorise patterns, leading to stronger correlations.

5 Clinical Analysis

To highlight the clinical utility of ML_ALM values, we undertook sex-stratified cross-sectional age-adjusted logistic regression to compare the associations of ground-truth ALM and predicted ML_ALM with low grip strength in the data used for cross-validation (N=30,483) as well as in a test dataset (N=21,687) from the same cohort without ground-truth ALM. The dataset represents a clinically relevant population (52% female, 65.2±7.8 years, 170.1±9.4 cm, 75.5±15.2 kg, 26.0±4.4 kg/m^2). Low ALM is associated with muscle weakness, which is a major driver of falls, functional decline, frailty, and sarcopenia. Hence, grip strength is clinically important and widely available in this and many other datasets. We did not evaluate hard clinical outcomes such as falls directly due to the comparably young age of the cohort with low number of incident falls (<2%).

Table 2: Odds ratios (OR) with 95% confidence intervals (CI) for the association between ALM and ML_ALM with low grip strength.

Dataset	Group	Measure	OR (95% CI) per SD increase
Cross-validation	Male	Prevalence (%)	1031 (7.2)
		ALM	0.53 (0.50, 0.57)
		ML_ALM	0.60 (0.55, 0.64)
	Female	Prevalence (%)	1083 (6.7)
		ALM	0.55 (0.51, 0.59)
		ML_ALM	0.61 (0.57, 0.66)
Test	Male	Prevalence (%)	733 (6.7)
		ML_ALM	0.70 (0.64, 0.77)
	Female	Prevalence (%)	881 (8.2)
		ML_ALM	0.81 (0.75, 0.88)

Standard deviations (SD) used for scaling predictors: Cross-validation: male ALM 3.06 kg, ML_ALM 2.74 kg; female ALM 2.24 kg, ML_ALM 2.04 kg. Test: male ML_ALM 3.46 kg; female ML_ALM 2.59 kg; low grip strength: male <27 kg, female <16 kg

The odds ratios displayed in Table 2 show that each sex-specific standard deviation increase in ALM and ML_ALM is associated with reduced odds for low grip strength. The association between ground-truth ALM and ML_ALM was similar in the cross-validation dataset. Notably, the relationship between ML_ALM and low grip strength in the test dataset was still significant, albeit weaker, which might be attributed to different anthropometric characteristics in this sub-set.

6 Conclusion

In this study we developed a multimodal deep learning model to estimate ALM from routinely captured DXA images of the hip and AP spine and clinical data. The proposed model incorporates comparably light-weight image encoders and fusion mechanisms to capture information from all modalities. Extensive experiments were undertaken to investigate the importance of single features for the prediction of ALM and to find the most expressive fusion mechanism.

Prospects of Application: The clinical analysis highlights that this tool could be used to flag individuals with low ALM who are more likely to be at risk of weakness during routine osteoporosis screening. Early identification of such individuals from routinely obtained DXA images serves as a form of opportunistic screening of those who might be at risk for falls and fractures. However, further external and cross-scanner validation is needed.

Acknowledgments. The salary of MM is supported by the Raine Medical Research Foundation through Raine Priming Grant (RPG045-2025). JRL is supported by a National Heart Foundation of Australia Future Leader Fellowship (102817; 107323). SZG is supported by the Western Australian Future Health Research and Innovation Fund (WANMAIdeas2025-26). NCH is supported by the UK Medical Research Council (MCPC21003; MCPC21001), the National Institute for Health Research (NIHR305844), and through the NIHR Southampton Biomedical Research Centre (NIHR203319). Salary of AKG (WANMAEL2025-26/8) and MS (WANMAEL2024-25/21) is supported by the Western Australian Future Health Research and Innovation Fund.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ballesteros-Pomar, M.D., González-Arnáiz, E., Pintor-de-la Maza, B., Barajas-Galindo, D., Ariadel-Cobo, D., González-Roza, L., Cano-Rodríguez, I.: Bioelectrical impedance analysis as an alternative to dual-energy X-ray absorptiometry in the assessment of fat mass and appendicular lean mass in patients with obesity. Nutrition **93**, 111442 (2022). https://doi.org/10.1016/j.nut.2021.111442
2. Beaudart, C., Alcazar, J., Aprahamian, I., Batsis, J.A., Yamada, Y., Prado, C.M., Reginster, J.Y., Sanchez-Rodriguez, D., Lim, W.S., Sim, M., von Haehling, S., Woo, J., Duque, G., in Sarcopenia (GLIS) group, G.L.I.: Health outcomes of sarcopenia: a consensus report by the outcome working group of the Global Leadership Initiative in Sarcopenia (GLIS). Aging Clin Exp Res **37**(1), 100 (Mar 2025). https://doi.org/10.1007/s40520-025-02995-9
3. 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., Topinková, E., Vandewoude, M., Zamboni, M., on Sarcopenia in Older People, E.W.G.: Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing **39**(4), 412–423 (Jul 2010). https://doi.org/10.1093/ageing/afq034

4. Cruz-Jentoft, A.J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., Sayer, A.A., Schneider, S.M., Sieber, C.C., Topinkova, E., Vandewoude, M., Visser, M., Zamboni, M.: Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing **48**(1), 16–31 (2019). https://doi.org/10.1093/ageing/afy169
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848
6. Feng, Q., Bešević, J., Conroy, M., Omiyale, W., Lacey, B., Allen, N.: Comparison of body composition measures assessed by bioelectrical impedance analysis versus dual-energy X-ray absorptiometry in the UK Biobank. Clin Nutr ESPEN **63**, 214–225 (Oct 2024). https://doi.org/10.1016/j.clnesp.2024.06.040
7. González-Martin, A.M., Limón-Villegas, E.S., Reyes-Castillo, Z., Esparza-Ros, F., Hernández-Palma, L.A., Santillán-Rivera, M.S., Herrera-Amante, C.A., Ramos-García, C.O., Righini, N.: Anthropometric Measurements for Predicting Low Appendicular Lean Mass Index for the Diagnosis of Sarcopenia: A Machine Learning Model. J Funct Morphol Kinesiol **10**(3), 276 (2025). https://doi.org/10.3390/jfmk10030276
8. Guo, W., Wang, J., Wang, S.: Deep Multimodal Representation Learning: A Survey. IEEE Access **7**, 63373–63394 (2019)
9. Huang, S.C., Pareek, A., Seyyedi, S., et al.: Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med **3**, 136 (2020). https://doi.org/10.1038/s41746-020-00341-z
10. Littlejohns, T.J., Holliday, J., Gibson, L.M., et al.: The uk biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nat Commun **11**, 2624 (2020). https://doi.org/10.1038/s41467-020-15948-9
11. Maqsood, A., Saleem, A., Sim, M., Suter, D., Radavelli-Bagatini, S., Hodgson, J.M., Prince, R.L., Zhu, K., Leslie, W.D., Schousboe, J.T., Lewis, J.R., Gilani, S.Z.: From Pixels to Prognosis: A Multi-Modal Attention-based Framework for Visceral Adipose Tissue Estimation. In: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) (2025)
12. Olshvang, D., Harris, C., Chellappa, R., Santhanam, P.: Predictive modeling of lean body mass, appendicular lean mass, and appendicular skeletal muscle mass using machine learning techniques: A comprehensive analysis utilizing NHANES data and the Look AHEAD study. PLoS One **19**(9), e0309830 (Sep 2024). https://doi.org/10.1371/journal.pone.0309830
13. Perez, E., Strub, F., de Vries, H., Dumoulin, V., Courville, A.: FiLM: Visual Reasoning with a General Conditioning Layer. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2018)
14. Rossum, K., Alexiuk, M.R., Bohm, C., Leslie, W.D., Tangri, N.: Development of an Algorithm to Predict Appendicular Lean Mass Index From Regional Spine and Hip DXA Scans. J Clin Densitom **28**(2), 101560 (2025). https://doi.org/10.1016/j.jocd.2024.101560
15. Ryu, J., Eom, S., Kim, H.C., Kim, C.O., Rhee, Y., You, S.C., Hong, N.: Chest X-ray-based opportunistic screening of sarcopenia using deep learning. J Cachexia Sarcopenia Muscle **14**(1), 418–428 (2023)
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks (2019)

17. de Santana Correia, A., Colombini, E.L.: Attention, please! A survey of neural attention models in deep learning. Artif Intell Rev **55**(8), 6037–6124 (2022). https://doi.org/10.1007/s10462-022-10148-x
18. Shi, J., He, Q., Pan, Y., Zhang, X., Li, M., Chen, S.: Estimation of appendicular skeletal muscle mass for women aged 60–70 years using a machine learning approach. J Am Med Dir Assoc **23** (2022). https://doi.org/10.1016/j.jamda.2022.09.002
19. Takahashi, S., Sakaguchi, Y., Kouno, N., et al.: Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. J Med Syst **48**, 84 (2024). https://doi.org/10.1007/s10916-024-02105-8
20. Thackeray, M., Orford, N.R., Kotowicz, M.A., Mohebbi, M., Pasco, J.A.: Estimation of whole-body and appendicular lean mass from spine and hip dual energy x-ray absorptiometry: A cross-sectional study. Calcified Tissue International **110**(3), 341–348 (Mar 2022). https://doi.org/10.1007/s00223-021-00922-4, epub 2021 Oct 13
21. World Health Organization: Ageing and health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (2023), accessed: 2026-01-30
22. Zhou, Y., Chen, S., Wang, Y., Huan, W.: Review of research on lightweight convolutional neural networks. In: 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). pp. 1713–1720 (2020). https://doi.org/10.1109/ITOEC49072.2020.9141847