Sarcopenia Assessment Model Based on Dual-Source Modal Graph

Wenxian Zheng[1], Zhi Chen[1], Qiaoqin Li[1], Rongyao Hu[1,#], Yongguo Liu[1,#]

[1]School of Information and Software Engineering, University of Electronic Science and Technology of China
zheng4wx@hotmail.com, zhic_cn@163.com, helenli803@163.com, hurongyao123@gmail.com, liuyg_cn@163.com

Abstract

Accurate muscle-mass assessment is crucial for staging and managing sarcopenia, yet existing methods suffer from modality-specific limitations and weak integration of muscle function indicators. To solve these limitations, we propose a Dual-source Features Graph for Sarcopenia Evaluation (DFGSE) to synergize high- and low-energy whole-body Dual-energy X-ray Absorptiometry (DXA) images, local high-energy DXA images, and blood-borne biochemical markers. Specifically, the feature extraction module employs dual-energy feature extraction to disentangle soft-tissue and skeletal cues from low-energy images, while skeleton-aware detection extracts joint features from high-energy images. It yields global and local DXA embeddings, complemented by blood-test representations. In the relevance exploration module, inter- and intra-modality correlations are computed via bilinear transformations to form adjacency matrices for the global, local, and blood modality representations. These matrices seed the Multi-type Multi-relation Graph Convolutional Network (MMGCN) – the core of the relation learning module – which captures both direct and indirect interactions among modalities through relation-aware message passing. Finally, the graph-fused representations are used by a muscle-mass prediction head trained with cross-entropy loss. Experiments on the public MURA dataset and two independent sarcopenia cohorts demonstrate that DFGSE consistently outperforms machine learning and state-of-the-art graph-based methods, in terms of four evaluation metrics for classification task.

Introduction

Sarcopenia is a progressive, age-related geriatric syndrome characterized by a decrease in muscle mass, strength, and physical performance. It can lead to a decline in the quality of life of the elderly. In the diagnosis of sarcopenia, muscle mass loss is a core element. Accurate assessment of muscle mass is conducive to the staging diagnosis of sarcopenia.

According to the sarcopenia guidelines and relevant expert consensus (Cruz-Jentoft A. J. et al, 2019), DXA (Dualenergy X-ray Absorptiometry) and BIA (Bioelectrical Impedance Analysis) are used to conduct muscle mass assessment. DXA is an imaging technique that uses low-dose X-rays at two energy levels to measure bone mineral density and body composition, and BIA is a method that estimates body composition by measuring the resistance of bodily tissues to a small electrical current. However, there are several limitations can be found, (1) These two methods only measure lean body mass, which is only a component of muscle mass (Chianca V. et al, 2022). (2) Muscle itself is in a hydrated state, which affects the diagnostic accuracy of the test because these two methods have poor ability to distinguish between extracellular fluid and muscle, leading an overestimation of muscle mass/quantity in individuals with excessive body fluids (Damluji A. A. et al, 2023), especially for patients with heart failure. (3) DXA and BIA are only applicable to muscle mass, and have a weak correlation with other muscle strength and physical function indicators (Manini T. M. et al, 2012). To solve them, Clark et al. (2014) proposed to use blood indicators, such as methyl-D3 creatine, to evaluate muscle mass. The operation of blood indicators is relatively simple, as well as is easy to conduct interpretability for muscle strength and physical function indicators. Therefore, it is necessary to combine DXA and blood indicators to diagnose muscle mass, which can improve the accuracy and interpretability of muscle mass evaluation.

Currently, most models for muscle mass evaluation use DXA and blood indicators (Moroni A et al, 2023; Zupo R et al, 2023). Usually, some blood indicators, such as serum 25(OH) vitamin D, are selected to measure the body muscle mass of a person, and the lean tissue information of the body is obtained by combining DXA images. Machine learning models based on DT, SVM, LASSO, etc. are used to identify some features in data that are most relevant to muscle mass.

Many traditional machine learning methods, such as linear regression and logistic regression, rely on linear assumptions. This may not be ideal when dealing with complex nonlinear relationships, and muscle mass assessment often involves complex physiological and biological nonlinear relationships (Moroni A et al, 2023; Tuttle C. S. L. et al, 2020; Hu et al, 2025). Muscle mass assessment may involve many different types of data, such as DXA images and blood indicators. When integrating these heterogene-

[#]Corresponding authors

ous data, traditional machine learning models have difficulty in simultaneously processing the structured information of blood test results and the unstructured information of DXA images, resulting in insufficient model performance (Hu et al, 2025). The models using DXA do not process the features contained in high-energy and low-energy X-rays separately, nor discuss the relationship between high-energy and low-energy rays, resulting in the failure to fully utilize the tissue features contained in low-energy ray images and the skeleton features contained in high-energy ray images, resulting in insufficient model accuracy.

The use of deep learning methods can solve the above problems, such as MPNet (Yao et al, 2023), KEGLN (Zhu et al, 2024) and Cascade-UDA(Zhan et al, 2025), and can also perform dynamic quality assessment based on data, thereby enabling early detection and diagnosis, and personalized intervention and prognosis after diagnosis. We get ever further, that a dual-source features graph in sarcopenia evaluation model (DFGSE) is proposed here. The overall architecture of the model is shown in Figure 1. The model consists of three main modules: (1) feature extraction module; (2) relevance exploration module; (3) relation learning module. First, the modal feature extraction module obtains human muscle mass and distribution characteristics from DXA images and blood biochemical indicators respectively, allowing different encoders to represent the corresponding modalities. Secondly, by treating different encoding modes as vertex-level features of the multimodal graph, modal correlation is used to measure the relationship between these vertices. Finally, graph convolution is introduced to achieve complementary modal relationship construction, and all relation graph are used in predicting whether a muscle mass loss event occurs.

To sum up, there are our contribution: improved ways to get the features to diagnose sarcopenia, especially on DXA features; propose a relation construct and learning method to get a significant accuracy on sarcopenia patient's diagnoses.

Approach

In this section, we introduce the proposed method - Dual-source Features Graph for Sarcopenia Evaluation, or DFGSE in short - with three parts, i.e., feature extraction, relevance exploration and relation learning module. The overall architecture can be found in Figure 1.

This study focuses on how to use existing DXA images and blood test results to evaluate and predict muscle mass. It is expected to predict whether the patient's muscle mass is reduced by inputting a whole-body high-energy DXA image $\boldsymbol{I_b^1}$ of $L_b \times H_b$ pixels, a whole-body low-energy DXA image $\boldsymbol{I_b^2}$ of $L_b \times H_b$ pixels, a local high-energy

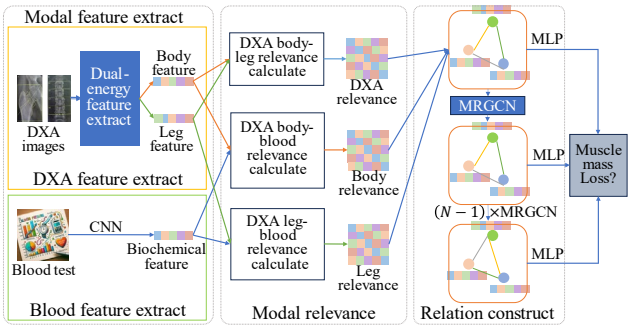

Figure 1: The architecture of Dual-source Features Graph in Sarcopenia Evaluation model (DFGSE).

DXA image $\boldsymbol{I_l}$ of $L_l \times H_l$ pixels, and a blood test report $\boldsymbol{b} \in \mathbb{R}^{1\times d}$. In order to solve the problem of poor muscle mass prediction caused by incomplete use of DXA and blood test features in traditional methods, this study proposes a corresponding solution model, DFGSE.

Feature Extraction Module. For X-ray images and blood test features extraction, MPNet (Yao et al, 2023) use attention-based X-ray image feature extraction module to get X-ray feature, and use CNN blood test feature extraction module to get blood feature. However, MPNet is designed to use diagnose pediatric pneumonia, so it only uses high-energy X-ray images of chest. To predict muscle condition and diagnose sarcopenia, both high-energy and low-energy X-ray images are need. So we improve MPNet, and propose our own feature extraction module.

In the feature extraction module, which is meanly shown in Figure 2, two important features for sarcopenia diagnosis are introduced: DXA images and blood biochemical measurement results. DXA uses two kinds of energies of X-rays (a high energy beam and a low energy beam).

In the DXA feature extraction module, it is necessary to input the whole body DXA image $\boldsymbol{I_b} = \left(\boldsymbol{I_{b,0}^1}, \boldsymbol{I_{b,0}^2}\right) \in \mathbb{R}^{2\times L_b^0 \times H_b^0}$ and the leg DXA image $\boldsymbol{I_l} \in \mathbb{R}^{L_l^0 \times H_l^0}$. The whole body DXA feature $\boldsymbol{f_b^{d,o}}$ is obtained by extracting features from the high-energy image and low-energy image of DXA (dual-energy feature extraction). Since the image produced by low-energy irradiation of leg DXA in the diagnosis of sarcopenia has poor muscle differentiation, only the high-energy ray image of the leg is used to extract the leg DXA feature $\boldsymbol{f_l^{d,o}}$.

When diagnosing sarcopenia, hospitals use dual-energy X-rays of the whole body and local high-energy X-ray images mainly of the hip. Therefore, these three images are used here for DXA feature extraction. For the whole body high- and low-energy X-ray images, the images are divided according to the human body boundaries to obtain the human body X-ray image and other background images. The high-energy X-ray image $\boldsymbol{I_b^1} \in \mathbb{R}^{L_b \times H_b}$ and low-energy X-ray image $\boldsymbol{I_b^2} \in \mathbb{R}^{L_b \times H_b}$ of the human body are retained,

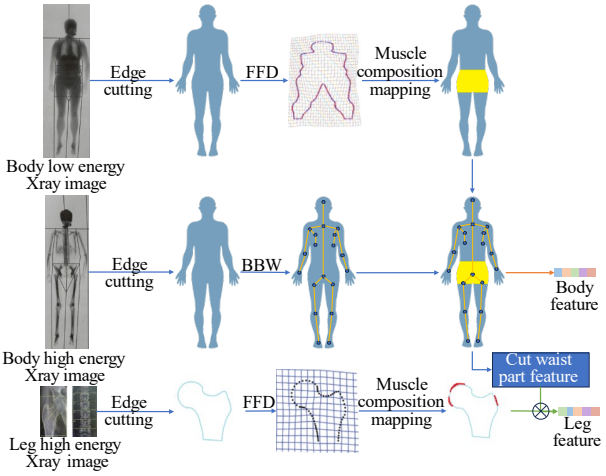

Figure 2: Dual-energy feature extraction.

and the background image is removed to improve the information density of the DXA image for subsequent processing.

$$\boldsymbol{I_b^1}, \boldsymbol{I_b^2} = \text{Shape}(\boldsymbol{I_b}) \tag{1}$$

Since high-energy images and low-energy X-ray images reveal different parts of the human body to different de grees, different methods are adopted here to extract the distribution characteristics of the muscles themselves and the skeleton extension features respectively. Low-energy X-ray images have a stronger ability to reveal human muscle tissue, so this image is used for muscle feature extraction. Here, FFD (Free-Form Deformation) is used to convert the divided human body shape into a human 3D model $\boldsymbol{M_b} \in \mathbb{R}^{L_b \times H_b \times C_b}$, and then the generated 3D model is further projected and transformed (orthogonal projection is used here) to obtain the muscle distribution characteristics $\boldsymbol{f_b^p} \in \mathbb{R}^{L_b \times H_b}$ under the standard human model.

$$\boldsymbol{M_b} = \text{FFD}(\boldsymbol{I_b^1}) \tag{2}$$

$$\boldsymbol{f_b^p} = \text{Project}(\boldsymbol{M_b}) \tag{3}$$

High-energy X-ray images have a stronger ability to reveal the human skeleton. Therefore, through this image, we can extract and capture the key parts of the human body, so as to obtain the muscle distribution information of specific parts, so as to compare with the muscle information obtained from the local image and integrate the local features from the global. Here, the BBW skeleton recognition framework is used to identify the body joints and obtain the skeleton feature map $\boldsymbol{g_b} \in \mathbb{R}^{L_b \times H_b}$.

$$\boldsymbol{g_b} = \text{BBW}(\boldsymbol{I_b^2}) \tag{4}$$

After obtaining the muscle distribution characteristics of the human body and the skeleton, the two are combined to obtain the global muscle features $\boldsymbol{f_b^{d,o}} \in \mathbb{R}^{L_b \times H_b}$ and the local muscle features $\boldsymbol{f_l^{b,o}} \in \mathbb{R}^{L_l \times H_l}$ of the corresponding skeleton parts.

$$\boldsymbol{f_b^{d,o}} = \text{CNN}(\boldsymbol{f_b^p w^p} \oplus \boldsymbol{g_b}) \tag{5}$$

$$\boldsymbol{f_l^{b,o}} = \text{F}(\boldsymbol{f_b^d}) \tag{6}$$

Where $\boldsymbol{w^p}$ is the feature projection weight, which can correspond the feature dimension of the muscle feature $\boldsymbol{f_b^d}$ to the feature dimension of the skeleton, thereby realizing the combination of the two features. $\text{F}(\cdot)$ is the feature interception function, which can intercept the muscle distribution characteristics of the specified area for reference in the processing results of the local high-energy X-ray image area.

For local high-energy X-ray images, the image is divided into local areas according to the local edge contour of the human body, and the local image of the human body and the background part are obtained. Only the human body part image $I_l^t \in \mathbb{R}^{L_l \times H_l}$ is retained, and the background image is removed to improve the information density of the DXA image for subsequent processing.

$$\boldsymbol{I_l^t} = \text{Shape}(\boldsymbol{I_l}) \tag{7}$$

Corresponding to the local high-energy X-ray, it emphasizes the local bone features, through which the corresponding skeletal muscle features can be inferred to obtain the local muscle features. Here, the FFD free shape transformation is also used to convert the divided local bone shape into a 3D model $\boldsymbol{M_l} \in \mathbb{R}^{L_l \times H_l}$, and then the generated 3D model is further projected and transformed (orthogonal projection is used here), and feature inference is performed to obtain the distribution characteristics of the muscle under the standard model $\boldsymbol{f_l^{p,o}} \in \mathbb{R}^{L_l \times H_l}$.

$$\boldsymbol{M_l} = \text{FFD}(\boldsymbol{I_l^t}) \tag{8}$$

$$\boldsymbol{f_l^{p,o}} = \text{Project}(\boldsymbol{M_l}) \tag{9}$$

After obtaining the local muscle features and the global muscle mass features of the corresponding parts, the two are fused to obtain the final local muscle features $\boldsymbol{f_l^d} = \boldsymbol{f_l^b} \otimes \boldsymbol{f_l^p} \in \mathbb{R}^{L_l \times H_l}$, where $\otimes$ is the Hadamard product.

The blood feature extraction module inputs the test results of all blood index tests performed during the examination $\boldsymbol{b} \in \mathbb{R}^{1 \times d}$, and obtains the whole-body average muscle biochemical characteristics $\boldsymbol{f^b} \in \mathbb{R}^{1 \times d}$ through blood feature extraction.

$$\boldsymbol{f^b} = \text{CNN}(\boldsymbol{b}) \tag{10}$$

In the process of evaluating muscle quality in real-life sarcopenia, dynamic data from multiple measurements may be included. Since the DXA measurement method has a certain degree of radioactivity, it is impossible to perform multiple measurements, resulting in a lack of matching DXA data for dynamic muscle blood test data. Therefore, this paper considers that the model can simulate the characteristics corresponding to the DXA modality based on dynamic blood test data, so that dynamic modeling of muscle quality can be performed.

The features of DXA modality samples (DXA global and DXA local) are combined with the time-series muscle blood test features. The obtained DXA sample features can be divided into DXA measurement features at different

time points, solving the problem that a single modality has time series while other modalities has only a single sample.

It is assumed that for the initial time point, DXA modality data and blood test data can be obtained at the same time; for subsequent time points, only muscle blood test data can be obtained but DXA data cannot be obtained.

For time point $t \geq 1$, the attention mechanism is used to calculate the enhanced scores α_i, β_i of the bleeding detection features:

$$\alpha_t = (\boldsymbol{w}^a)^\top \tanh\left(\boldsymbol{W}^a\left(\boldsymbol{f}_{\boldsymbol{d_b}} + a_t \boldsymbol{f}^{\boldsymbol{b_t}}\right) + \boldsymbol{b}^a\right) \quad (11)$$

$$\beta_t = (\boldsymbol{w}^b)^\top \tanh\left(\boldsymbol{W}^b\left(\boldsymbol{f}_{\boldsymbol{d_b}} + b_t \boldsymbol{f}^{\boldsymbol{b_t}}\right) + \boldsymbol{b}^b\right) \quad (12)$$

Where, $\boldsymbol{w}^a, \boldsymbol{w}^b$ are attention weights, $\boldsymbol{W}^a, \boldsymbol{W}^b$ are feature learning weights, $\boldsymbol{b}^a, \boldsymbol{b}^b$ are feature learning biases, and a_t, b_t are the enhanced biases that determine the amount of blood test features included in the enhanced features.

According to the enhancement scores calculated in the above way, the DXA global features and DXA local features are enhanced to obtain the enhanced features $\boldsymbol{f}'_{\boldsymbol{b_i}}, \boldsymbol{f}''_{\boldsymbol{b_i}}$ of the corresponding modality:

$$\boldsymbol{f}'_{\boldsymbol{b_t}} = \alpha_t \boldsymbol{f}^{\boldsymbol{b_t}}, \boldsymbol{f}''_{\boldsymbol{b_t}} = \beta_t \boldsymbol{f}^{\boldsymbol{b_t}} \quad (13)$$

Finally, the original modality features are fused by enhanced features to obtain time-series enhanced time-series DXA data. The enhanced features of DXA global and DXA local adopt different fusion methods. Both DXA global modality and muscle blood test modality focuses on the information of muscles throughout the body, so these two can be treated equally, that is, combined by exponential moving average summation, λ_t^α is used to enhance the fusion weights of DXA whole-body modalities.

$$\boldsymbol{f}^{\boldsymbol{d}}_{\boldsymbol{b_t}} = (1 - \lambda_t^\alpha)\boldsymbol{f}^{\boldsymbol{d,o}}_{\boldsymbol{b}} + \lambda_t^\alpha \boldsymbol{f}'_{\boldsymbol{b_t}} \quad (14)$$

The DXA local features focus more on the information of local muscles, while the whole-body muscle information can only provide a certain reference for the local information and cannot completely replace the local information itself. Therefore, we adopted a weighted approach to combine them, λ_t^β is used to enhance the fusion weights of DXA leg modalities.

$$\boldsymbol{f}^{\boldsymbol{d}}_{\boldsymbol{l_t}} = \boldsymbol{f}^{\boldsymbol{d,o}}_{\boldsymbol{l}} + \lambda_t^\beta \boldsymbol{f}''_{\boldsymbol{b_t}} \quad (15)$$

Then integrating the enhanced DXA data at each time point, the time-series enhanced muscle mass features are obtained: $f_b^d = \left(f_{b_1}^d | f_{b_2}^d | \cdots | f_{b_T}^d\right), f_l^d = \left(f_{l_1}^d | f_{l_2}^d | \cdots | f_{l_T}^d\right)$, where $|$ is the concatenation operation.

Relevance Exploration Module. The global feature $\boldsymbol{f}^{\boldsymbol{d}}_{\boldsymbol{b}}$ and leg feature $\boldsymbol{f}^{\boldsymbol{d}}_{\boldsymbol{l}}$ from DXA, as well as the global muscle average feature $\boldsymbol{f}^{\boldsymbol{b}}$ from blood indicators, are obtained from the modal feature extraction module. At this time, in order to combine the features of different modalities (DXA and blood test), it is necessary to measure the feature correlation of different modalities, so as to better construct the correlation diagram of different modalities and then select a better strategy for fusion.

The DXA modality can be decomposed into two parts,

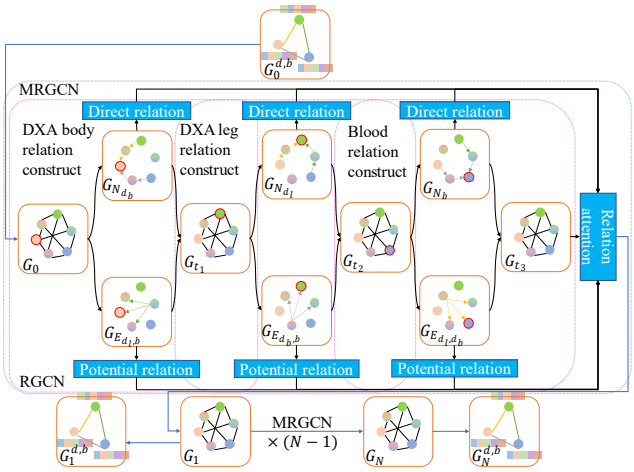

Figure 3: Modal relationship construction: MMGCN.

representing the global muscle characteristics and the leg muscle characteristics. By measuring the correlation between these two parts, the blood characteristics can better divide the global muscle characteristics and the local muscle characteristics, and the fusion of the average muscle characteristics can better represent the characteristics of different levels, so as to better perform the diagnosis work related to muscle quality.

Here, the blood feature $\boldsymbol{f}^{\boldsymbol{b}}$ is measured for global correlation with the DXA global feature $\boldsymbol{f}^{\boldsymbol{d}}_{\boldsymbol{b}}$ to obtain the overall correlation of muscle mass, and the muscle mass global correlation matrix $\boldsymbol{A_b} \in \mathbb{R}^{L_b \times H_b}$ is constructed; and the blood feature $\boldsymbol{f}^{\boldsymbol{b}}$ is measured for local correlation with the DXA local feature $\boldsymbol{f}^{\boldsymbol{d}}_{\boldsymbol{l}}$ to obtain the muscle mass correlation matrix $\boldsymbol{A_l} \in \mathbb{R}^{L_l \times H_l}$ in the leg. Through these two correlation matrices, the associated multimodal joint information can be obtained from the global level and the leg level, respectively, so that the modal feature map can be constructed later for optimization learning.

$$\boldsymbol{A_b} = \sigma\left(f_b^d W_b^d f^b W^b\right), \boldsymbol{A_l} = \sigma\left(f_l^d W_l^d f^b W^b\right) \quad (16)$$

After obtaining the inter-modality correlation matrix, normalization is performed to obtain $\overline{\boldsymbol{A}}_{\boldsymbol{b}}, \overline{\boldsymbol{A}}_{\boldsymbol{l}}$.Similarly, the intra-modality correlation $\boldsymbol{A^d} \in \mathbb{R}^{L_b \times H_b}$ is calculated for the two parts of the DXA modality - the global muscle feature and the local muscle feature.

$$\boldsymbol{A^d} = \sigma\left(f_b^d W_d^b f_l^d W_l^b\right) \quad (17)$$

After normalizing the intra-modal correlation matrix, we get $\overline{\boldsymbol{A}}^{\boldsymbol{d}}$.

Relation Learning Module. Disease diagnose hardly combines multiple modals. For example, KEGLN (Zhu et al, 2024) only use electronic medical record to construct patient's disease relation graph. To fit the sarcopenia patients' diagnose, we design a new approach to combine DXA image and blood test result to get a better result on muscle mass prediction.

The multimodal graph $\boldsymbol{G^{d,b}}$ of muscle mass can be con-

structed by measuring the modal correlation. For the multimodal graph whose adjacency matrix is measured by bilinear transformation, the use of graph convolution can be further applied to model deep graph patterns, which helps to explore the inter-modal relationships and obtain complementary feature representations with less redundancy.

We introduce a Multimodal Relational Graph Convolutional Network (MMGCN) to integrate three feature types—full-body DXA muscle data, localized DXA muscle data, and biochemical indicators. MMGCN consists of three Relational Graph Convolutional (RGCN) modules—one per modality—and a relational attention mechanism. Each RGCN models both direct influences (as edges between modalities) and indirect influences (mediated by the third modality), producing an intermediate relationship graph after each pass. Once all three RGCNs have generated their graphs, the relational attention mechanism fuses direct and indirect representations to capture the true intermodal interactions. Repeating this convolution process N times yields a final graph that accurately reflects the complex interplay among whole-body muscle, local muscle, and biochemical features.

Within each RGCN model, the input relationship graph is split into a direct graph—linking the target modality directly to the other two—and an indirect graph—encoding paths through the remaining modality. By merging these two views, the network fully leverages multimodal correlations and improves its ability to assess muscle quality.

$$\boldsymbol{G_{t_i}} \rightarrow \boldsymbol{G_v^n}, \boldsymbol{G_e^n} \rightarrow \boldsymbol{G_{t_{i+1}}} \tag{18}$$

The above RGCN module can be used to obtain the relationship graphs for the three modalities of DXA whole body, DXA local and muscle biochemical indicators. However, the focus of the relationship graph learning obtained by RGCN is still on the transformation of the three modalities, and does not consider the learning of their relationship transformation. In drug association acquisition, it is necessary to obtain relationship changes. The RGCN module can obtain direct and indirect relationships, and the direct and indirect relationships are vectorized through relationship learning to obtain the corresponding relationship representations for the three modalities.

For DXA whole body modality:

$$\boldsymbol{r_d^{(1)}} = \mathrm{RL}\big(G_v^{d_b}, R_d\big), \boldsymbol{r_p^{(1)}} = \mathrm{RL}\big(G_e^{d_b}, R_p\big) \tag{19}$$

For DXA local modality:

$$\boldsymbol{r_d^{(2)}} = \mathrm{RL}\big(G_v^{d_l}, R_d\big), \boldsymbol{r_p^{(2)}} = \mathrm{RL}\big(G_e^{d_l}, R_p\big) \tag{20}$$

For Muscle biochemical index modality:

$$\boldsymbol{r_d^{(3)}} = \mathrm{RL}(G_v^b, R_d), \boldsymbol{r_p^{(3)}} = \mathrm{RL}\big(G_e^b, R_p\big) \tag{21}$$

After obtaining the direct and indirect representations of the three modalities, the relationship representations of each modality are spliced together to obtain a unified direct representation and a unified indirect representation in the entire muscle modality relationship diagram:

$$\boldsymbol{r_d} = \boldsymbol{r_d^{(1)}} || \boldsymbol{r_d^{(2)}} || \boldsymbol{r_d^{(3)}}, \boldsymbol{r_p} = \boldsymbol{r_p^{(1)}} || \boldsymbol{r_p^{(2)}} || \boldsymbol{r_p^{(3)}} \tag{22}$$

After obtaining the relationship evolution of nodes and the changes in the relationship itself, it is necessary to cognitively aggregate the nodes and relationships to complete the learning of the relationship graph. Here, the relationship-aware attention mechanism is used for processing. First, the attention scores of direct and indirect relationships are obtained:

$$a_d^{(i)} = (\boldsymbol{w}^{a,d})^\top \tanh\big(\boldsymbol{W}^{a,d} r_d^{(i)} + \boldsymbol{b}^{a,d}\big) \tag{23}$$

$$a_p^{(i)} = (\boldsymbol{w}^{a,p})^\top \tanh\big(\boldsymbol{W}^{a,p} r_p^{(i)} + \boldsymbol{b}^{a,p}\big) \tag{24}$$

$$\alpha_d^{(i)} = \frac{\exp a_d^{(i)}}{\sum_{t'=1}^{3} \exp a_d^{(t')}}, \alpha_p^{(i)} = \frac{\exp a_p^{(i)}}{\sum_{t'=1}^{3} \exp a_p^{(t')}} \tag{25}$$

Then, the relationship features are weighted and fused through the attention score to obtain a relationship graph neural network containing direct and indirect relationships, where $\boldsymbol{\alpha_d} = \big(\alpha_d^{(1)}, \alpha_d^{(2)}, \alpha_d^{(3)}\big), \boldsymbol{\alpha_p} = \big(\alpha_p^{(1)}, \alpha_p^{(2)}, \alpha_p^{(3)}\big)$:

$$\boldsymbol{G_n} = \mathrm{Fusion}\big(\boldsymbol{G_{t_3}}, \boldsymbol{r_d}, \boldsymbol{r_p}, \boldsymbol{\alpha_d}, \boldsymbol{\alpha_p}\big) \tag{26}$$

Since each layer of the relationship graph contains the relationship information between each modality to a certain extent when constructing the relationship, fusing all the relationship graphs here can better predict whether the muscle content has decreased.

For all relation graph $\boldsymbol{G}_n$, we can get a prediction result $\hat{y}_n$ shows that if the patient has the muscle mass loss issue, by using multilayer perceptron, which is known as MLP(·).

$$\hat{y}_n, \hat{p}_n = \mathrm{MLP}(\boldsymbol{G}_n) \tag{27}$$

$\hat{p}_k$ is the confidence rate for the prediction result.

The final fused modality muscle mass relationship diagram can be obtained from the modality relationship fusion module, and the final muscle mass prediction can be made through this relationship diagram.

$$\hat{y} = \underset{\hat{y}_n}{\operatorname{argmax}}\, \hat{p}_n \tag{28}$$

The cross-entropy loss function is used for training in muscle mass prediction.

$$\ell = -\sum_n [y_n \log \hat{y}_n + (1 - y_n) \log(1 - \hat{y}_n)] \tag{29}$$

For all N layers MMGCN, we use relevant patient's test n-th muscle loss result as label y_n. To get the best performer of the DFGSE model, the minimum loss is required.

Experiment

Datasets. The experimental datasets use the public MURA dataset and real-life sarcopenia patient cohort data from Sichuan Provincial People's Hospital (SAMSPH) and the Mianyang 404 Hospital (SCMY404). More details can be found in the supplementary materials.

Comparison Methods. To compare with the proposed method, we employ three kinds of methods: (1) baseline (Hsu et al, 2023): Logistic regression, support vector machine (SVM), XGBoost, KNN, and random forest (RF)

MURA	ACC	F1	SEN	SPE
Baseline	80.32±2.31	65.24±1.75	57.97±1.62	**90.39±2.67**
GA-Net	80.94±1.93	79.63±1.54	73.68±1.83	75.84±1.68
GeCoM	81.06±1.76	76.19±1.63	68.37±1.58	85.62±1.83
HIT	82.69±2.02	83.93±2.00	82.33±1.94	85.68±1.79
GFN	82.17±1.84	84.64±1.97	81.84±1.75	88.29±1.84
R-HGNN	83.06±2.05	82.70±1.86	81.73±1.68	86.07±1.62
ERGNN	82.38±1.87	80.76±1.58	75.90±1.47	85.02±1.83
DFGSE	**84.38±1.93**	**85.18±1.46**	**83.07±2.00**	89.16±1.82

Table 1: Classification results (%) of all methods with four evaluation metrics in MURA dataset (average±std).

SAMSPH	ACC	F1	SEN	SPE
Baseline	84.22±2.94	69.82±1.38	62.48±0.99	**91.82±2.03**
GA-Net	81.57±1.65	76.14±1.82	74.03±1.20	76.82±1.62
GeCoM	84.94±1.76	78.69±1.69	67.33±1.06	87.13±1.68
HIT	85.04±1.97	80.47±1.84	79.65±1.54	82.60±1.70
GFN	83.64±1.89	84.51±1.70	**84.68±1.36**	87.57±1.84
R-HGNN	84.87±1.93	83.64±1.81	82.47±1.24	90.27±1.83
ERGNN	83.43±1.29	84.17±1.27	79.58±1.35	88.21±1.64
DFGSE	**86.22±1.83**	**85.01±1.29**	83.79±1.63	88.10±1.58

Table 2: Classification results (%) of all methods with four evaluation metrics in SAMSPH dataset (average±std).

		ACC	F1	SEN	SPE
R_l ($R_p = 0.3$)	0.2	84.67±1.30	83.94±1.03	80.86±1.67	87.63±0.99
	0.4	**86.22±1.83**	**85.01±1.29**	**83.79±1.63**	**88.10±1.58**
	0.6	85.91±1.68	84.27±1.35	83.27±1.92	87.83±1.76
	0.8	82.94±1.52	82.06±1.25	81.03±1.20	84.93±0.98
R_p ($R_l = 0.4$)	0.1	81.87±1.62	80.99±0.93	79.87±1.34	82.78±0.86
	0.2	84.39±1.60	83.75±1.08	81.76±1.28	84.91±1.35
	0.3	**86.22±1.83**	**85.01±1.29**	**83.79±1.63**	**88.10±1.58**
	0.4	85.16±1.59	84.14±1.28	82.43±1.50	86.83±1.29

Table 3: Classification results (%) of different DFGSE parameters in SAMSPH dataset (average±std).

models are used here. (2), neural network models: GA-Net (Dai et al, 2024), GeCoM (Wang et al, 2024), HIT (Keller et al, 2024). (3) graph neural network: GFN (Hu el al, 2021), R-HGNN (Yu et al, 2022), ERGNN (Li et al, 2025). We use the original settings on their paper. More introduction is shown in the Comparison Methods part in supplementary material.

Settings. The experiments run on computing platform which has Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz processor, NVIDIA GeForce RTX 3090 24GB graphic card, and 400GB RAM.

To evaluate the performance of the DFGSE model, the following commonly used classification evaluation indicators are selected: accuracy (ACC), F1 score (F1), sensitivity (SEN), specificity (SPE). We divide the samples into 80%, 10%, and 10% for training, verification, and testing in MURA dataset. Then, the real-world datasets (SAMSPH and SCMY404) are used for testing. The optimization strategy of the DFGSE model is Adam gradient descent, and the cross-entropy function is selected for back propagation. The optimal parameters of the model are set to a learning rate of 0.001, a weight decay of 0.15, and a random dropout rate of 0.3. For comparison methods, we follow the settings in their original paper.

Experimental Results. The results can be found in Table 1 and Table 2. Through experiments in the MURA dataset and a real-life sarcopenia cohort, the DFGSE model obtains the best performance for those datasets and followed

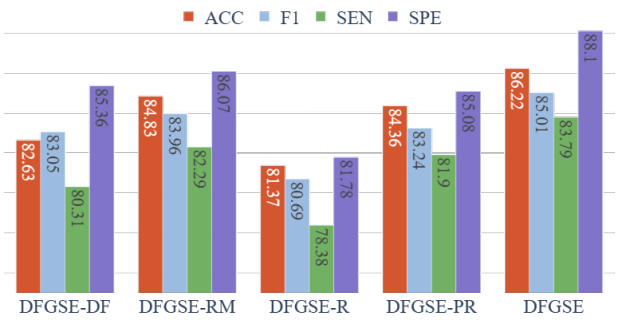

Figure 4: Classification results (%) of DFGSE model ablation on SAMSPH dataset.

by GFN, ERGNN, R-HGNN, HIT, GeCoM, baseline and GA-Net. Compared with the best comparison method GFN, the average improvement in ACC, F1, SEN, SPE is 2.40, 0.52, 0.17 and 0.70 percentage points respectively. Compared with the worst comparison method GA-Net, the average improvement in ACC, F1, SEN and SPE is 4.05, 7.21, 9.68 and 9.30 percentage points respectively.

As the baseline, classical machine learning methods get the best SPE result, its SPE is average above DFGSE 2.29 percentage points. However, for some indicators (F1, SEN), the baseline methods get the worse classification results, the average improvement in F1 and SEN is 17.57 and 20.53 percentage points respectively, and in ACC we got average improvement is 3.03 percentage points. They only get the features from all inputs, not considering all relationship almost all modal inputs. For neural network methods (GA-Net, GeCoM, HIT), the average improvement in ACC, F1, SEN, SPE is 2.76, 5.92, 9.20 and 6.35 percentage points respectively. They improved their ways to extract the modal features, but they still not consider the relations for all these modals. For graph neural network methods (GFN, R-HGNN, ERGNN), the average improvement in ACC, F1, SEN, SPE is 2.04, 1.69, 2.40 and 1.06 percentage points respectively. They consider the relations on or even behind all modals, but they get the relationship in indirect way. DFGSE can get the best result, by separating all modals relation in direct and potential relation on a more notable way.

Ablation study. We conduct an ablation study to evaluate the effectiveness of key Duel components by comparing three variants with the full model (Figure 4): (1) DFGSE-DF: Remove DXA dual feature extraction module. (2) DFGSE-RM: Remove the multimodal relationship exploration part of DXA whole-body, DXA local and biochemical indicators. (3) DFGSE-R: Remove whole relationship learning module. (4) DFGSE-PR: Remove modal relation graph's potential relationship learning part.

DFGSE gets the best result and is followed by DFGSE-RM, DFGSE-PR, DFGSE-R, DFGSE-DF. Comparing with DFGSE-R, DFGSE can get the maximum improvement in ACC, F1, SEN and SPE, which is 4.85, 4.34, 6.41, 6.32 percentage points respectively. The experimental results show every module that the module with the greatest contribution in the DFGSE model is relation learning module. Relation learning module strengthens the model's feature perception of muscle quality during the learning process by performing explicit relation learning on muscle mass-related modalities, thereby greatly improving the model's indicator effect. Both direct relationship construction and indirect relationship construction play important role, which shows that it is necessary to incorporate indirect relation into the relation learning network. The dual feature extraction module of the model makes full use of the information of DXA images and blood test features, so that the indicators of the model can be further improved. The introduction of the correlation matrix makes the model more interpretable when constructing the relationship graph, and improves the accuracy of the model in muscle mass assessment to a certain extent.

Parameter Study. In the process of constructing the graph, two parameters are involved, $\boldsymbol{R_l}$ for direct relation extraction and $\boldsymbol{R_p}$ for indirect relation representation. In order to explore the optimal values of the above parameters, the parameter experiments shown in Table 3 were conducted. From the experimental result, both $\boldsymbol{R_l}$ and $\boldsymbol{R_p}$ follow the same trend: first rising and then falling.

It shows that R_l achieves the best effect at 0.4, which can maximize the inclusion of the mode-mode direct relationship in the constructed relationship diagram. Comparing with the worst effect result, it improves 3.3 percentage points. R_p achieves the best effect at 0.3, which improves maximum percentage points up to 4.4, that can maximize the use of the correlation between the other two modes and the influence of the remaining modes when constructing the relationship diagram, thereby improving the performance of the model.

Conclusion

This paper designed DFGSE to addresses the limitations of existing sarcopenia assessment methods by integrating dual-energy DXA images and blood biomarkers into a unified graph-based model. The framework captured soft-tissue and skeletal features, simulated longitudinal DXA trends from dynamic blood tests, and explored high-order inter-modal relationships. Extensive experiments on the public MURA dataset and two real-world sarcopenia cohorts demonstrate DFGSE's out performs and state-of-the-art methods across accuracy, F1-score, sensitivity and specificity metrics.

Future work will explore incorporating additional functional indicators and deploying the model in clinical workflows for real-time, personalized sarcopenia management.

References

Chianca V.; Albano D.; Messina C.; Gitto S.; Ruffo G.; Guarino S.; Grande F. D. and Sconfienza L. M. 2022. Sarcopenia: imaging assessment and clinical application. *Abdominal Radiology*, *47*(9), 3205-3216.

Damluji A. A.; Alfaraidhy M.; AlHajri N.; Rohant N. N.; Kumar M., Malouf C. A.; Bahrainy S.; Kwak M. J.; Batchelor W. B.; Forman D. E.; Rich M. W.; Kirkpatrick J.; Krishnaswami A.; Alexander K. P.; Gerstenblith G.; Cawthon P.; deFilippi C. R. and Goyal P. 2023. Sarcopenia and cardiovascular diseases. *Circulation*, 147(20): 1534-1553.

Manini T. M.; Clark B. C. 2012. Dynapenia and aging: an update. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 67(1), 28-40.

Clark R. V.; Walker A. C.; O'Connor-Semmes R. L.; Leonard M. S.; Miller R. R.; Stimpon S. A.; Turner S. M.; Ravussin E.; Cefalu W. T.; Hellerstein M. K. and Evans W.. 2014. Total body skeletal muscle mass: estimation by creatine (methyl-d3) dilution in humans. *Journal of Applied Physiology*, 116(12), 1605-1613.

Moroni A.; Perna S.; Azzolino D.; Gasparri C.; Zupo R.; Cremasco M. M. and Rondanelli M. 2023. Discovering the individualized factors associated with sarcopenia and sarcopenic obesity phenotypes—a machine learning approach. *Nutrients*, 15(21), 4536.

Zupo R.; Moroni A.; Castellana F.; Gasparri C.; Catino F.; Lampignano L.; Perna S.; Clodoveo M. L.; Sardone R. and Rondanelli M. 2023. A machine-learning approach to target clinical and biological features associated with sarcopenia: findings from northern and southern Italian aging populations. *Metabolites*, 13(4), 565.

Hu R.; Zhan M.; Gan J.; Li L.; Ye F.; and Liu T. 2025. Resilient kernel-based unsupervised multi-view feature selection via compact binary hashing. *Engineering Applications of Artificial Intelligence*, 162, 112515.

Cruz-Jentoft A. J.; Bahat G.; Bauer J.; Boirie Y.; Bruyere O.; Cederholm T; Copper C.; Landi F.; Rolland Y.; Sayer A. A.; Schneider S. M.; Sieber C. C.; Topinkova E.; Vandewoude M.; Visser M.; Zamboni M.; members of European Working Group on Sarcopenia in Older People 2 (EWGSOP2), and the Extended Group for EWGSOP2. 2019. Sarcopenia: revised European consensus on definition and diagnosis. *Age and ageing*, 48(1), 16-31.

Tuttle C. S. L.; Thang L. A. N.; Maier A. B. 2020. Markers of inflammation and their association with muscle strength and mass: A systematic review and meta-analysis. *Ageing research reviews*, 64(1), 101185.

Hu R.; Gan J.; Zhan M.; Li L.; and Wei M. 2025. Unsupervised Kernel-based Multi-view Feature Selection with Robust Self-representation and Binary Hashing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(16), 17287-17294.

Zhan M.; Wu Z.; Xu H.; Zhu X.; and Hu R. 2025. Cascade-UDA: A Cascade paradigm for unsupervised domain adaptation. Neurocomputing, 636, 129924.

Hsu W. H.; Ko A. T.; Weng C. S.; Chang C. L.; Jan Y. T.; Lin J. B.; Chien H. J.; Lin W. C.; Sun F. J.; Wu K. Pin. and Lee J. 2023. Explainable machine learning model for predicting skeletal muscle loss during surgery and adjuvant chemotherapy in ovarian cancer. *Journal of Cachexia, Sarcopenia and Muscle*, 14(5), 2044-2053.

Dai J.; Liu T.; Torigian D. A.; Tong Y.; Han S.; Nie P.; Zhang J.; Li R.; Xie F. and Udupa J. K. 2024. Ga-net: a geographical attention neural network for the segmentation of body torso tissue composition. *Medical image analysis*, 91: 102987.

Hu K.; Wang Z.; Martens K. A. E.; Hagenbuchner M.; Bennamoun M. and Tsoi A. 2021. Graph fusion network-based multimodal learning for freezing of gait detection. *IEEE Transactions on Neural Networks and Learning Systems*, 34(3), 1588-1600.

Yu L.; Sun L.; Du B.; Liu C.; Lv W. and Xiong H. 2022. Heterogeneous graph representation learning with relation awareness. *IEEE Transactions on Knowledge and Data Engineering*, 35(6), 5935-5947.

Li G.; Yang J. and Liang S. 2025. ERGNN: Spectral graph neural network with explicitly-optimized rational graph filters. In Processing of *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.1-5.

Wang Y.; Zhen L.; Tan T. E.; Fu H.; Feng Y. and Wang Z. 2024. Geometric correspondence-based multimodal learning for ophthalmic image analysis. *IEEE Transactions on Medical Imaging*, 43(5), 1945-1957.

Keller M.; Arora V.; Dakri A.; Chandhok S.; Machann J. and Fritsche A. 2024. HIT: Estimating internal human implicit tissues from the body surface. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3480-3490.

Rajpurkar P.; Irvin J.; Bagul A.; Ding D.; Duan T.; Mehta H.; Yang B.; Zhu K.; Laird D.; Ball R. L.; Langlotz C.; Shapanskaya K.; Lungren M. P.; Ng, A. Y. 2017. Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv preprint*, arXiv:1712.06957.

Yao D.; Xu Z.; Lin Y.; and Zhan Y. 2023. Accurate and intelligent diagnosis of pediatric pneumonia using X-ray images and blood testing data. *Frontiers in bioengineering and biotechnology*, 11, 1058888.

Zhu Z.; Li J.; Xu C.; Zou J.; and Zhao Q. 2024. A knowledge-guided event-relation graph learning network for patient similarity with Chinese electronic medical records. *IEEE Transactions on Big Data*, 11(3), 1475-1492.