From Pixels to Prognosis: A Multi-Modal Attention-based Framework for Visceral Adipose Tissue Estimation

Arooba Maqsood[1,2](✉), Afsah Saleem[1,2], Marc Sim[1,2,3], David Suter[1,2], Simone Radavelli-Bagatini[2], Jonathan M. Hodgson[2,3], Richard L. Prince[3], Kun Zhu[3], William D. Leslie[5], John T. Schousboe[6], Joshua R. Lewis[1,2], and Syed Zulqarnain Gilani[1,2,4]

[1] Centre for AI & ML, School of Science, Edith Cowan University, Australia
a.maqsood@ecu.edu.au
[2] Nutrition & Health Innovation Research Institute, Edith Cowan University, Australia
[3] Medical School, University of Western Australia, Australia
[4] Department of Computer Science, University of Western Australia, Australia
[5] Department of Medicine & Radiology, University of Manitoba, Canada
[6] Park Nicollet Clinic & HealthPartners Institute, HealthPartners, Minneapolis, USA

Abstract. Obesity is a chronic disease that increases the risk of multi-organ damage as well as cardiovascular disease, diabetes, and certain cancers. It is strongly related to Visceral Adipose Tissue (VAT), which is the fat stored around the internal organs. New approaches to assessing VAT in large populations are essential to understand how obesity contributes to chronic disease progression. Various direct and indirect measures have been developed to quantify VAT. However, many of these techniques either fail to distinguish between various types of body fats (e.g., subcutaneous versus visceral) or involve high radiation imaging and/or are costly (e.g., Computed Tomography). Annually, millions of individuals globally undergo hip or spine Dual-energy X-ray Absorptiometry (DXA) scans to screen for osteoporosis as well as lateral spine (LS) scans to detect vertebral fractures. In this paper, we develop a multi-modal attention-based framework for VAT estimation from LS DXA scans and patient demographic information. We compare our results on two LS DXA datasets with baseline methods and also perform clinical analysis to demonstrate its effectiveness. The proposed approach has the potential to enable cost-effective, non-invasive, and efficient quantification of VAT in people undergoing bone density assessment with LS scans. To the best of our knowledge, this is the first paper to predict VAT from DXA LS scans.

Keywords: Obesity · Visceral Adipose Tissue · Lateral Spine DXA scans · Multi-modality · Feature Fusion

1 Introduction

Obesity, characterized by an excessive accumulation of body fat, poses a serious threat to global health [20]. It is a leading cause of morbidity and mortality worldwide, contributing to a wide range of chronic conditions including cardiovascular disease, and type-2 diabetes [4], placing a substantial burden on healthcare [1]. Obesity is associated with changes in adipose tissue, which is categorized into two main types based on their distinct locations and metabolic characteristics: (i) Subcutaneous Adipose Tissue (SAT) and (ii) Visceral Adipose Tissue (VAT). Although both are vital for various body functions, visceral adiposity is a major contributor to a range of serious health conditions, including cardiovascular disease, obesity-related cancers, high blood pressure, impaired glucose regulation, high triglycerides, and low high-density lipoprotein cholesterol [2,4,17].

Obesity is often measured using indirect methods such as body mass index, waist circumference, waist-to-height ratio and waist-to-hip ratio [30]. However, these methods do not capture the fat distribution or distinguish between VAT and SAT [7,30]. This oversimplification contributes to inconsistencies in diagnosis, highlighting the need for a more precise approach to measure obesity [7]. Advancements in medical imaging have paved the way for precise non-invasive methods for assessing VAT [29]. Imaging techniques such as digital X-rays, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) are used to evaluate fat mass [30,18]. However, X-rays and CT scans expose patients to moderate/high levels of ionizing radiation [15], while MRI is expensive [18]. Dual-energy X-ray Absorptiometry (DXA), on the other hand, is a cost-effective technique with minimal radiation exposure [15]. It is commonly used to evaluate bone density during routine osteoporosis screening [26,27]. Although whole-body DXA scans can measure VAT in terms of volume, mass, and android-to-gynoid fat mass ratios [19], they are not routinely obtained in clinical practice. Alternatively, DXA-derived Lateral Spine (LS) scans are often obtained in routine clinical practice to assess vertebral fractures [26] and offer the advantage of significantly shorter scan times. In this work, we leverage LS DXA scans instead of whole-body DXA scans and design a novel framework for VAT prediction. To the best of our knowledge, this is the first such attempt.

Several attempts have been made to quantify different types of adipose tissue using CT or MRI scans. Park et al. [21] utilized a Convolutional Neural Network (CNN) for automated segmentation of abdominal muscle and fat areas on CT scans. Feng et al. [9] proposed the *FM-Net* consisting of two Res-UNet blocks to localize Epicardial Adipose Tissue (EAT) from MRI scans. Qu et al. [22] also employed a variant of UNet for EAT segmentation using CT scans. Schneider et al. [29] used CNN models, including UNet [25], DenseUNet [3], and CDFNet [8], on MRI scans for abdominal fat quantification. It is worth noting that all these methods are based on segmentation techniques. However, obtaining manual annotations for VAT is a laborious and subjective process [13,29]. Another limitation of these studies is that they can be prone to inaccuracies in segmentation boundaries, leading to incorrect measurement of VAT. This paper aims to overcome these limitations by automating VAT measurement using DXA scans, a widely used imaging modality known for its low radiation exposure and cost.

Despite its benefits, DXA imaging faces challenges such as low signal-to-noise ratio, low contrast and image artifacts [15,10]. Furthermore, DXA scans of the lateral spine do not capture the full abdominal region. These limitations make it difficult to distinguish VAT from surrounding tissues and organs. To overcome this, we add patient demographic data routinely collected during DXA scanning for VAT prediction. This integration is expected to improve the accuracy of the VAT quantification when combined with DXA imaging data. However, fusing these modalities is complex due to the differences in data representations: demographic features such as age, weight, and height are structured numerical data, while DXA scans are unstructured, high-dimensional images. Aligning these modalities requires a robust fusion strategy to ensure meaningful interaction between the two.

To address this, we design a unified attention-based approach for VAT prediction from LS DXA scans. This approach ensures the extraction of modality-specific features while using the attention mechanism for the optimal integration of patient demographic and imaging data, improving model performance. Our experiments show that this added information complements the DXA images by providing additional context. A unique feature of our approach is that we quantify VAT from LS DXA scans, using labels derived from whole-body DXA scans. To further validate the relevance of our work, we conduct clinical analysis to investigate the impact of increased VAT on metabolic syndrome, which is strongly associated with obesity [11]. The code is available at: our GitHub repository.

The major contributions of this paper are as follows:

- To the best of our knowledge, this is the first study to automate VAT quantification from lateral spine DXA scans instead of whole-body DXA scans.
- We propose a novel attention-based approach for predicting VAT from LS DXA scans, integrating both imaging and demographic data to improve prediction accuracy.
- We validate our findings by examining the correlation of both actual and predicted VAT with specific clinical markers in 837 women. To highlight, in the test set of 784 women with clinical measures but no VAT measures and 694 women without diabetes but with measures of metabolic syndrome, our predicted VAT is strongly associated with both clinical markers and odds of having metabolic syndrome.

2 Methodology

The proposed unified approach, shown in Figure 1, consists of four main components: (i) a CNN-based feature extractor, (ii) a tabular encoder to encode demographic data, (iii) an attention-based fusion mechanism to integrate multimodal embeddings, and (iv) prediction of VAT using regression.

The proposed framework comprises a CNN-based encoder for image data and a Transformer-based encoder for demographic data due to the distinct strengths each model offers for their respective input types. CNNs excel at capturing spatial hierarchies and local patterns, making them ideal for image feature extraction. However, tabular data such as age, weight, and height lack spatial relationships, which limits the effectiveness of CNNs [31]. In contrast, Transformers excel at handling sequential or tabular data by capturing feature dependencies without relying on spatial structure [5,31].

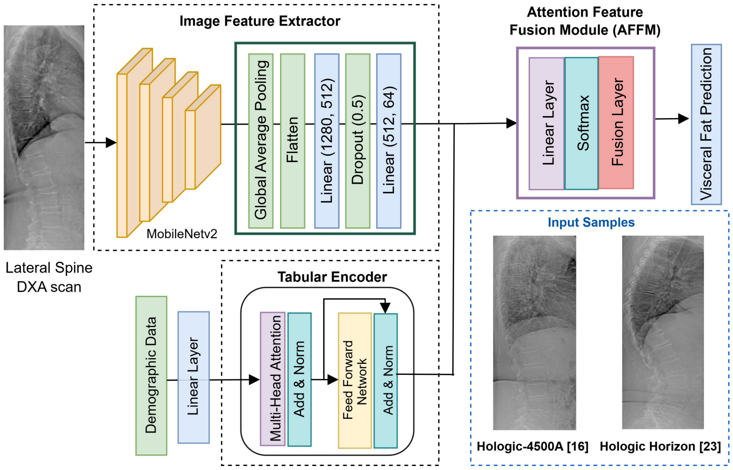


Fig. 1. Our proposed multi-modal attention-based model for VAT estimation.

Image Feature Extractor: Given an input image $X \in \mathbb{R}^{C\times H\times W}$, where C represents the number of channels, and H and W are the height and width of the image, we extract image embeddings using the Global Average Pooling (GAP) layer of a pretrained MobileNetv2 [28], which was trained on the ImageNet [6] dataset. The output of the GAP layer is a feature vector $X'_{img} \in \mathbb{R}^{1280}$. This feature map is then down-sampled via a custom fully connected layer, producing a 64-dimensional image embedding ($X'_{img} \in \mathbb{R}^{1280} \rightarrow Z_{\text{img}} \in \mathbb{R}^{64}$) which effectively captures the essential features of the image.

Tabular Encoder: Inspired by the vanilla transformer [34], we adapt and modify the encoder to learn tabular embeddings for patient demographics. Passing the numeric demographic variables through this block, we map them into a higher-dimensional embedding space, which allows the tabular encoder to capture complex relationships between the variables. The attention mechanism within the encoder allows the model to focus on the most relevant relationship [14], enhancing its capacity to learn complex patterns in the data. The demographic data for each scan is a vector of three features (see Eq. 1). This data is first converted into an embedding using a fully-connected layer.

$$D'_{\text{demographic}} = \left[\text{age}, \text{weight}, \text{height}\right] \in \mathbb{R}^3 \tag{1}$$

Positional embeddings were introduced in the transformer encoder [34] to learn contextual and positional information within the input sequence. Unlike text or time-series data, where positional relationships carry meaning, the order of features in our tabular dataset is arbitrary and does not impact the semantic interpretation. Hence, we replace the positional embedding layer with a linear projection, allowing the tabular encoder to focus on meaningful feature interactions rather than sequence order.

The tabular encoder processes these input embeddings using multi-head self-attention [34] and a point-wise feed-forward network [34]. While self-attention captures feature relationships, the feed-forward network applies non-linear transformations to refine feature representations. Multi-head self-attention computes attention scores (Eq. 2) using multiple heads, allowing the model to focus on different parts of the input sequence in parallel. This enables each head to capture distinct aspects of the input features.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{2}$$

The final output of the tabular encoder is pooled across the sequence length to generate a 64-dimensional demographic embedding ($Z_{\text{demographic}} \in \mathbb{R}^{64}$).

Attention Feature Fusion Module (AFFM): The AFFM block is designed to fuse the image embeddings (Z_{img}) and demographic embeddings ($Z_{\text{demographic}}$) effectively. By fusing the imaging and demographic data, we exploit the strengths of both modalities, capturing crucial visual and demographic patterns. The AFFM block consists of three main components: a linear layer, a softmax layer and a fusion layer. Firstly, image and demographic embeddings are independently extracted and passed through separate linear layers to compute modality-specific attention weights. After Softmax normalization, these weights are applied to the embeddings to perform weighted fusion. The attention weights α_{img} and $\alpha_{\text{demographic}}$ can be calculated as $\alpha_{\text{img}} = \text{softmax}(W_{\text{img}} \cdot Z_{\text{img}})$ and $\alpha_{\text{demographic}} = \text{softmax}(W_{\text{demographic}} \cdot Z_{\text{demographic}})$ respectively, where W_{img} and $W_{\text{demographic}}$ are the learnable attention weights.

These attention weights adaptively learn to assign more importance to the most relevant features, ensuring that crucial information is prioritized during the training process. Finally, the Fusion Layer applies these learned attention weights to the image and demographic embeddings, generating a single fused representation, effectively integrating both modalities.

$$f_{\text{fusion}} = \alpha_{\text{img}} \cdot Z_{\text{img}} + \alpha_{\text{demographic}} \cdot Z_{\text{demographic}} \tag{3}$$

The fused embeddings (f_{fusion}) are passed through a fully connected network comprising two hidden layers with ReLU activation and dropout regularization to prevent overfitting.

3 Experiments and Results

Datasets. We evaluate the effectiveness of our proposed model using two distinct datasets of LS DXA scans. The Hologic-4500A dataset acquired using the Hologic 4500A machine contains 2,285 single-energy lateral-spine DXA scans, with each scan having dimensions of 800 × 287 obtained from a Western Australian study of community dwelling ambulant women over the age of 70 years, the Perth Longitudinal Study of Aging in Women (PLSAW) [16]. The Hologic Horizon dataset contains a total of 466 scans from the Hologic Horizon machine, obtained from 245 community dwelling ambulant men and women aged 60 to 80 [23,24]. These scans have variable dimensions. Both datasets contain VAT mass labels in grams (g) obtained from whole-body DXA scans, serving as the ground truth for training the model. The datasets also include imaging data and demographic variables (age, weight, and height) collected at scan time.

Implementation Details. The images from both datasets are resized to a fixed target size of 800 × 287 to ensure uniformity, while avoiding artificial borders that may hinder learning. Single-channel DXA scans are replicated to 3-channels to match pretrained model inputs. Pixel values and demographic variables are normalized prior training. Data augmentations (±10° rotation and horizontal flipping with p=0.3) were applied. Flipping was tested but found less effective. All the experiments are performed in PyTorch using stratified 10-fold cross-validation. In each fold, the dataset is divided into 80% for training, 10% for validation and 10% for testing, ensuring that every sample is used for testing exactly once. The final performance is reported as average across all folds. Models are re-initialized in each fold to prevent data leakage. Validation loss is used to adjust the learning rate dynamically. The network is fine-tuned with a low initial learning rate (1×10^{-3}) using the Adam optimizer, trained for 25 epochs with batch size 32, and learning rate is reduced on plateau. Training settings including no. of epochs and learning rate were selected via grid search within the cross-validation. The primary loss function is Root Mean Squared Error (RMSE), with Mean Absolute Percentage Error (MAPE) monitored as a secondary metric.

Evaluation Metrics. The performance of the proposed model is evaluated using the RMSE, MAPE and Pearson's correlation.

Baseline. The MobileNetV2 [28] model, trained solely on images with RMSE loss, serves as the baseline. Instead of direct regression, embeddings from MobileNetV2 are processed via a 3×3 convolution, dropout (p=0.5), and two fully connected layers (512→128→1) for prediction.

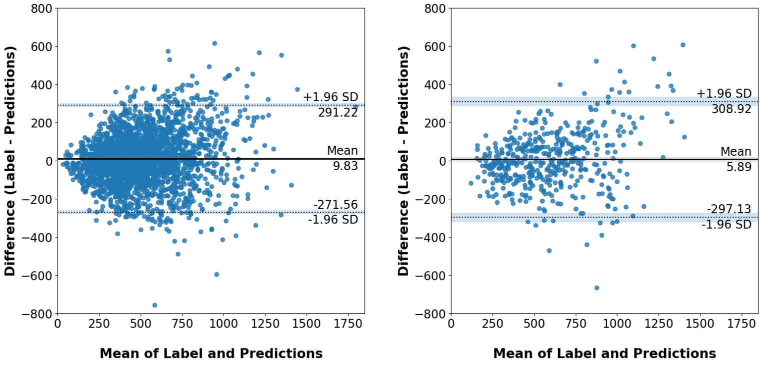
Results and Discussion. The results presented in Table 1 compare the performance of our proposed model with the baseline. Across both the Hologic-4500A and Hologic Horizon datasets, the proposed model consistently outperformed the baseline in terms of RMSE and MAPE. For the Hologic-4500A dataset, the MAPE reduced from 28.09% to 25.88%, while the RMSE reduced from 155.59g to 143.86g. The results demonstrate the ability of our model to generalize better than the baseline for VAT prediction, capturing patterns from the data more effectively. A similar trend is also observed in the Hologic Horizon dataset. Our model significantly improves with a MAPE reduction of 26.42% to 21.96% and RMSE reduction of 170.08g to 154.55g. Despite overlapping confidence intervals, paired t-tests confirmed the improvements were significant (p<0.001).

Table 1. Comparison of the proposed model with a baseline for VAT prediction using the Hologic-4500A [16] and Hologic Horizon [23,24] dataset.

Dataset	Model	MAPE(%) ↓	RMSE (g) ↓
Hologic-4500A [16]	Baseline	28.09 ±2.10	155.59 ±10.88
	Proposed	**25.88** ±0.75	**143.86** ±10.04
Hologic Horizon [24]	Baseline	26.42 ±2.90	170.08 ±13.50
	Proposed	**21.96** ±3.49	**154.55** ±13.89

The reduction in MAPE and RMSE across both datasets demonstrates the model's robustness and accuracy. These improvements stem from the model's multi-modal design, integrating imaging and demographic features to capture complex patterns.

The VAT predictions from the proposed model show a strong correlation with reference values from both the Hologic-4500A and Hologic Horizon datasets,



(a) Hologic-4500A [16] dataset (b) Hologic Horizon [23,24] dataset

Fig. 2. Bland-Altman plots. The plots show agreement between the actual values and predicted values against their mean for both understudy datasets.

with Pearson's correlation values of 0.828 and 0.846, respectively. Given that we have used the labels from whole-body DXA to train our model, the predictions are fairly accurate, considering the inherent challenges associated with LS DXA. Bland–Altman analysis (Figure 2) indicates minimal bias suggesting good agreement between the predictions and reference values, with mean differences of 9.83 (95% limits of agreement: -271.56, 291.22) and 5.89 (95% limits of agreement: -297.13, 308.92) for Hologic-4500A and Hologic Horizon respectively.

Ablation Studies. For the ablation study, we evaluate the impact of various CNN-based models and fusion techniques on the performance of our proposed model. First, we test VGG-16 [32], ResNet-18 [12], ResNet-50 [12], MobileNetv2 [28] and EfficientNetv2s [33] as feature extractors. Additionally, we conduct experiments using only demographic data to evaluate the impact of demographic features. We also replace the tabular encoder in the proposed model with a simple neural network-based encoder.

Table 2. Ablation study evaluating the performance of various CNN-based models and the impact of different techniques for encoding demographic features using the Hologic-4500A DXA scan dataset for VAT prediction.

Ablation Type	**Input**		**Model**	**MAPE(%) ↓**	**RMSE (g) ↓**
	Imaging Modality	**Tabular Modality**			
For image encoder	✓		VGG-16 [32]	42.79	228.81
	✓		ResNet-18 [12]	33.05	172.82
	✓		ResNet-50 [12]	31.97	163.80
	✓		MobileNetv2 [28]	28.09	155.59
	✓		EfficientNetv2s [33]	30.87	161.97
For tabular encoder		✓	Neural Network	50.54	249.52
		✓	Tabular Encoder	55.45	393.45
	✓	✓	Neural Network	29.32	164.38
	✓	✓	**Tabular Encoder**	**25.88**	**143.86**

The results presented in Table 2 clearly show that MobileNetv2 [28] outperforms the tabular encoder when used with an image feature extractor. Notably, the tabular encoder, with its attention mechanism, demonstrates superior performance compared to the neural network, highlighting its ability to capture

Table 3. Ablation study evaluating the impact of various feature fusion techniques on the performance of the proposed model using the Hologic-4500A [16] dataset.

Fusion Technique	**MAPE (%) ↓**	**RMSE (g) ↓**
Concatenation	27.93	178.39
Addition	29.06	184.13
Attention-based	**25.88**	**143.86**

complex and subtle inter-dependencies between demographic variables. Furthermore, our model shows significant improvement in VAT prediction, particularly when demographic information is incorporated. The ablation study provides strong evidence that combining demographic data with DXA imaging leads to enhanced accuracy in VAT prediction.

We also explore different feature-fusion strategies, including concatenation, addition, and our proposed attention-based fusion. As shown in Table 3, the attention-based fusion method outperforms by dynamically weighting the image and demographic data embeddings, leading to more accurate predictions.

Clinical Analysis: Obesity is a key driver of Metabolic Syndrome (Met-S) [11], with high triglyceride and low high-density lipoprotein cholesterol levels being major contributors [11]. To demonstrate the clinical significance of our work, we investigate the correlation between predicted VAT and demographic variables (i.e, age, height, weight), as well as markers of Met-S including total cholesterol (CHOL), Low-density Lipoprotein Cholesterol (LDL), High-density Lipoprotein Cholesterol (HDL), and Triglycerides (TRIG) in two cohorts of 1,404 people (referred to as D_{train}) where ground-truth VAT is available. To further test our model, we use 991 LS DXA images from one of the same cohorts captured 5 years earlier (referred to as D_{test}) where no ground truth for VAT is available.

Table 4. Spearman's correlation between VAT measures, age, height, weight, and circulating lipids.

VAT	**age**	**weight (Kg)**	**height (cm)**	**CHOL (mmol/L)**	**TRIG (mmol/L)**	**HDL (mmol/L)**	**LDL (mmol/L**
GT D_{train}	-0.079^{*}	0.720^{**}	0.033	-0.109^{**}	0.357^{**}	-0.346^{**}	-0.056
Pred D_{train}	-0.111^{**}	0.761^{**}	0.064	-0.123^{**}	0.336^{**}	-0.332^{**}	-0.072
Pred D_{test}	-0.099^{**}	0.661^{**}	0.052	-0.022	0.326^{**}	-0.327^{**}	0.011

** *p<0.01.* * *p<0.05*

First, we compute the correlation between the ground truth values of VAT (i.e., GT D_{train}) to establish a baseline for comparison (Row-1 Table 4). We then calculate the correlation between predictions of our model (i.e., Pred D_{train}) to assess whether these predictions align with the expected clinical trends. To validate our model, correlation analyses on previously unseen data (i.e., Pred D_{test}) show moderate to strong positive correlations between visceral fat, weight and TRIG, while HDL and CHOL exhibit negative correlations. Similar trends across cohorts confirm the consistency of associations.

We then exclude women with diabetes, or those missing data needed for Met-S diagnosis from D_{test} and apply age-adjusted logistic regression in 694 women to assess the odds of having Met-S by quartiles (Q) of predicted VAT. The proportion of women with Met-S was highest in Q4 (women with the highest predicted VAT) (53.6%), compared to Q3 (25.4%), Q2 (17.4%) and Q1 (12%). Compared to women in Q1, women in Q3 (OR 2.64 95%CI 1.49-4.68) and Q4

(OR 8.97 95%CI 5.16-15.61), but not Q2 (OR 1.56 95%CI 0.86-2.85), had significantly higher odds for Met-S. This demonstrates that predicted VAT may be a useful clinical tool when identifying those at risk of Met-S.

4 Conclusion

To our knowledge, this is the first automated method for predicting VAT from LS DXA images. Our multi-modal framework combines a CNN for image feature extraction and a tabular encoder with attention fusion to integrate demographic data. This framework was evaluated on two datasets and achieved state-of-the-art performance, with cross-sectional validation against Met-S in 684 older women. One limitation of this study is the lack of matching CT/MRI data, which prevented comparisons with other imaging modalities. Future work will focus on assessing the scalability of the model on larger datasets. Additionally, we plan to investigate vision transformers for image feature extraction and extend the model to predict both fat and muscle mass, expanding its applicability.

Acknowledgments and Data Use Declaration. De-identified labeled images for ML (Project# 03349-LEWIS) were sourced from existing VFA studies. Written informed consent was obtained from all participants involved. For the Hologic-4500A data, the Human Ethics Committee of the University of Western Australia approved the 5-year RCT and subsequent 10-year follow up (Approval# 05/06/004/H50). For Hologic-Horizon dataset, the HealthPartners Institutional Review Board [IRB] [#A20-149]) approved the use of the de-identified images and the ethics approval for the study was granted by the Edith Cowan University Human Research Ethics Committee (Project# 20513-HODGSON). The study was supported by a National Health and Medical Research Council (NHMRC) of Australia Ideas grant (APP1183570). SZG was partially funded by a Raine Priming Grant from the Raine Medical Research Foundation and a WANMA grant funded 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 content of this article.

References

1. https://www.worldobesity.org/about/about-obesity/prevalence-of-obesity
2. Bauer, C., Sim, M., Prince, R.L., Zhu, K., Lim, E.M., Byrnes, E., Pavlos, N., Lim, W.H., Wong, G., Lewis, J.R., et al.: Circulating lipocalin-2 and features of metabolic syndrome in community-dwelling older women: A cross-sectional study. Bone **176** (2023). https://doi.org/10.1016/j.bone.2023.116861
3. Cai, S., Tian, Y., Lui, H., Zeng, H., Wu, Y., Chen, G.: Dense-UNet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quantitative Imaging in Medicine and Surgery **10**(6) (2020)
4. Chang, S.H., Beason, T.S., Hunleth, J.M., Colditz, G.A.: A systematic review of body fat distribution and mortality in older people. Maturitas **72**(3) (2012). https://doi.org/10.1016/j.maturitas.2012.04.004

5. Choi, S.R., Lee, M.: Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. Biology **12**(7) (2023). https://doi.org/10.3390/biology12071033
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
7. Endocrinology, T.L.D.: Redefining obesity: Advancing care for better lives. https://doi.org/10.1016/S2213-8587(25)00004-X
8. Estrada, S., Conjeti, S., Ahmad, M., Navab, N., Reuter, M.: Competition vs. Concatenation in skip connections of fully convolutional networks. In: Machine Learning in Medical Imaging: 9th International Workshop. Springer (2018)
9. Feng, F., Carlhäll, C.J., Tan, Y., Agrawal, S., Lundberg, P., Bai, J., Yang, J.Z., Trew, M., Zhao, J.: FM-Net: A fully automatic deep learning pipeline for epicardial adipose tissue segmentation. In: International Workshop on Statistical Atlases and Computational Models of the Heart. Springer (2023). https://doi.org/10.1007/978-3-031-52448-6_9
10. Gilani, S.Z., Sharif, N., Suter, D., Schousboe, J.T., Reid, S., Leslie, W.D., Lewis, J.R.: Show, attend and detect: Towards fine-grained assessment of abdominal aortic calcification on vertebral fracture assessment scans. In: MICCAI (2022). https://doi.org/10.1007/978-3-031-16437-8_42
11. Han, T.S., Lean, M.E.: A clinical perspective of obesity, metabolic syndrome and cardiovascular disease. JRSM Cardiovascular Disease **5** (2016). https://doi.org/10.1177/2048004016633371
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)
13. Hemke, R., Buckless, C.G., Tsao, A., Wang, B., Torriani, M.: Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. Skeletal radiology **49** (2020). https://doi.org/10.1007/s00256-019-03289-8
14. Hernández, A., Amigó, J.M.: Attention mechanisms and their applications to complex systems. Entropy **23** (2021). https://doi.org/10.3390/e23030283
15. Ilyas, Z., Sharif, N., Schousboe, J.T., Lewis, J.R., Suter, D., Gilani, S.Z.: GuideNet: Learning inter-vertebral guides in DXA lateral spine images. In: DICTA (2021). https://doi.org/10.1109/DICTA52665.2021.9647067
16. Lewis, J.R., Schousboe, J.T., Lim, W.H., Wong, G., Wilson, K.E., Zhu, K., Thompson, P.L., Kiel, D.P., Prince, R.L.: Long-term atherosclerotic vascular disease risk and prognosis in elderly women with abdominal aortic calcification on lateral spine images captured during bone density testing: A prospective study. Journal of Bone and Mineral Research **33**(6) (2018). https://doi.org/10.1002/jbmr.3405
17. Lopes, H.F., Corrêa-Giannella, M.L., Consolim-Colombo, F.M., Egan, B.M.: Visceral adiposity syndrome. Diabetology & Metabolic Syndrome **8** (2016). https://doi.org/10.1186/s13098-016-0156-2
18. Maskarinec, G., Shvetsov, Y.B., Wong, M.C., Garber, A., Monroe, K., Ernst, T.M., Buchthal, S.D., Lim, U., Le Marchand, L., Heymsfield, S.B., et al.: Subcutaneous and visceral fat assessment by DXA and MRI in older adults and children. Obesity **30**(4) (2022). https://doi.org/10.1002/oby.23381
19. Mohammad, A., Ziyab, A.H., Mohammad, T.: Anthropometric and dxa-derived measures of body composition in relation to pre-diabetes among adults. BMJ Open Diabetes Research and Care **11**(5) (2023). https://doi.org/10.1136/bmjdrc-2023-003412

20. Obesity Evidence Hub: Disease burden of overweight, obesity and poor diet, https://www.obesityevidencehub.org.au/collections/impacts/disease-burden-overweight-obesity-poor-diet
21. Park, H.J., Shin, Y., Park, J., Kim, H., Lee, I.S., Seo, D.W., Huh, J., Lee, T.Y., Park, T., Lee, J., et al.: Development and validation of a deep learning system for segmentation of abdominal muscle and fat on Computed Tomography. Korean Journal of Radiology **21**(1) (2020). https://doi.org/10.3348/kjr.2019.0470
22. Qu, J., Chang, Y., Sun, L., Li, Y., Si, Q., Yang, M.F., Li, C., Zhang, X.: Deep learning-based approach for the automatic quantification of epicardial adipose tissue from non-contrast CT. Cognitive Computation **14**(4) (2022). https://doi.org/10.1007/s12559-022-10036-0
23. Radavelli-Bagatini, S., Bondonno, C.P., Dalla Via, J., Sim, M., Gebre, A.K., Blekkenhorst, L.C., Connolly, E.L., Bondonno, N.P., Schousboe, J.T., Woodman, R.J., et al.: Impact of provision of abdominal aortic calcification results on fruit and vegetable intake: 12-week randomized phase 2 controlled trial. Nature Communications **15**(1) (2024). https://doi.org/10.1038/s41467-024-52172-1
24. Radavelli-Bagatini, S., Bondonno, C.P., Sim, M., Blekkenhorst, L.C., Anokye, R., Connolly, E., Bondonno, N.P., Schousboe, J.T., Woodman, R.J., Zhu, K., et al.: Modification of Diet, Exercise and Lifestyle (MODEL) Study: A randomised controlled trial protocol. BMJ open (2020). https://doi.org/10.1136/bmjopen-2019-036366
25. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: (MICCAI). pp. 234–241. Springer (2015)
26. Rühling, S., Schwarting, J., Froelich, M.F., Löffler, M.T., Bodden, J., Hernandez Petzsche, M.R., Baum, T., Wostrack, M., Aftahy, A.K., Seifert-Klauss, V., et al.: Cost-effectiveness of opportunistic QCT-based osteoporosis screening for the prediction of incident vertebral fractures. Frontiers in Endocrinology **14** (2023). https://doi.org/10.3389/fendo.2023.1222041
27. Saleem, A., Ilyas, Z., Suter, D., Hassan, G.M., Reid, S., Schousboe, J.T., Prince, R., Leslie, W.D., Lewis, J.R., Gilani, S.Z.: Scol: Supervised contrastive ordinal loss for abdominal aortic calcification scoring on vertebral fracture assessment scans. In: MICCAI (2023). https://doi.org/10.48550/arXiv.2307.12006
28. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetv2: Inverted residuals and linear bottlenecks. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
29. Schneider, D., Eggebrecht, T., Linder, A., Linder, N., Schaudinn, A., Blüher, M., Denecke, T., Busse, H.: Abdominal fat quantification using convolutional networks. European Radiology **33**(12) (2023). https://doi.org/10.1007/s00330-023-09865-w
30. Shuster, A., Patlas, M., Pinthus, J., Mourtzakis, M.: The clinical importance of visceral adiposity: A critical review of methods for visceral adipose tissue analysis. The British Journal of Radiology **85**(1009) (2012). https://doi.org/10.1259/bjr/38447238
31. Shwartz-Ziv, R., Armon, A.: Tabular data: Deep learning is not all you need. Information Fusion **81** (2022). https://doi.org/10.48550/arXiv.2106.03253
32. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
33. Tan, M., Le, Q.: EfficientNetv2: Smaller models and faster training. In: International Conference on Machine Learning. PMLR (2021)
34. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)