VisiFAT: Estimating Visceral Fat from Dual-View DXA scans using Cross-View Attention

Arooba Maqsood[1,2], Abdullah Tariq[1], Marion Mundt[1,2], Marc Sim[1,2], David Suter[1,2], Nicholas C. Harvey[4,5], William D. Leslie[6], John T. Schousboe[7], Joshua R. Lewis[1,2], and Syed Zulqarnain Gilani[1,2,3]

[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] Department of Computer Science, University of Western Australia, Australia
[4] MRC Lifecourse Epidemiology Centre, University of Southampton, UK
[5] NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK
[6] Departments of Medicine & Radiology, University of Manitoba, Canada
[7] Park Nicollet Clinic & HealthPartners Institute, HealthPartners, Minneapolis, USA

Abstract. Visceral Adipose Tissue (VAT) is a key biomarker for cardiometabolic disease risk, but its accurate quantification typically requires Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), which are costly and not routinely available for large-scale or longitudinal assessment. Dual-energy X-ray Absorptiometry (DXA), on the other hand, offers a low-radiation, cost-effective alternative. However, existing research on VAT estimation from DXA remains limited and relies on single-view–based methods. In this work, we introduce *VisiFAT*, a multi-view, multi-modal framework for estimating VAT from lateral and posteroanterior spine DXA scans, together with patient demographic data including sex, age, height, and weight. The proposed approach employs a dual-branch convolutional architecture coupled with a cross-view attention module that enables adaptive feature exchange between views. By jointly modelling complementary DXA projections and patient-level information, the framework captures three-dimensional body composition cues that are not observable from either view alone. Evaluated on a large cohort, the proposed approach outperforms single-view and conventional fusion baselines, achieving lower prediction error and stronger correlation with reference VAT measures. To assess clinical relevance, we further examined cross-sectional associations between both estimated and reference VAT and the prevalence of atherosclerotic cardiovascular disease and diabetes, demonstrating comparable risk stratification performance. These results support *VisiFAT* as a scalable tool for VAT estimation and cardiometabolic risk assessment.

Keywords: Obesity · Visceral Adipose Tissue · Medical Imaging · Dual-energy X-ray Absorptiometry · Cross-View Attention · UK Biobank · Deep Learning · Multi-Modal Learning · Fat Quantification · Feature Fusion · Body Composition Analysis · Scalable AI

1 Introduction

Obesity is a chronic disease characterized by excess body fat and is a major contributor to global morbidity and mortality [15]. In 2022, around 1 in 8 people worldwide, or more than 1 billion people, were living with obesity, underscoring its status as the major public health challenge of the 21st century [1,18].

Obesity-related risk is not determined by total body fat but by how adipose tissue is distributed [15], with particular concern over Visceral Adipose Tissue (VAT). VAT refers to fat deposits located within the abdominal cavity and represents a highly metabolically active adipose compartment [11]. It is strongly associated with adverse health outcomes including insulin resistance, type-II diabetes, cardiovascular disease, and some cancers [3,6,15]. Accurate assessment of VAT may help improve cardiometabolic risk-stratification and clinical decision-making.

VAT can be estimated either indirectly using anthropometric measures or directly using imaging-based techniques including Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Dual-energy X-ray Absorptiometry (DXA). Indirect approaches include Body Mass Index (BMI), waist circumference, waist-to-hip ratio, and waist-to-height ratio. However, these measures are inherently limited and fail to distinguish visceral from subcutaneous fat [10,15]. Although BMI remains a useful population-level screening tool, it does not capture regional fat distribution or accurately reflect total body fat percentage. Moreover, the substantial variability in body composition across sex, age, and ethnicity further limits its clinical utility [2]. Collectively, these limitations underscore the need for more precise and clinically meaningful measures of adiposity. Reflecting this, recent international consensus statements have emphasized the importance of abdominal obesity [2,5].

Whilst imaging-based techniques overcome shortcomings of anthropometric measures, they have their own limitations. CT and MRI are considered reference standards for VAT estimation. However, both are time-consuming, CT exposes patients to significant doses of ionizing radiation [7], while MRI, although widely regarded as the gold standard, provides complementary measures such as fat fraction estimation, is expensive [11]. These factors can limit their scalability and routine use in large-scale or population level screening [11]. Although whole-body DXA scans are relatively inexpensive, have negligible ionizing radiation [7], and can estimate visceral fat mass and distribution, they are not routinely performed [2,10]. However, there is a growing recognition that low-cost, scalable [10], and widely accessible [5] approaches to VAT quantification should be integrated into the routine clinical workflow. To close this gap, we aim to use the spine DXA scans, that are already collected in people attending bone density testing for osteoporosis screening.

Research on opportunistic VAT estimation using routinely acquired DXA scans remains limited. Maqsood et al. [10] demonstrated that Lateral Spine (LS) DXA images, commonly acquired for vertebral fracture assessment as part of osteoporosis screening [10,12], can be leveraged for VAT estimation. This study provides early evidence supporting the use of routine LS scans as a scalable

solution for VAT estimation. However, this work implicitly assumes that a single projection sufficiently captures the anatomical cues associated with visceral adiposity. This assumption is restrictive, as visceral fat accumulates in three dimensions and is influenced by trunk geometry including abdominal depth and width, which may not be fully represented in a single view. Moreover, the two cohorts used in this study were all women or predominantly women [10], which limits generalizability as visceral adiposity is known to differ substantially between males and females [8,9].

As a part of osteoporosis screening, lateral and posteroanterior spine (PAS) scans can be obtained concurrently, with each projection likely to offer complementary anatomical information for VAT estimation, without performing explicit 3D reconstruction. Motivated by this, we reformulated opportunistic VAT estimation as a multi-view learning problem. While simple feature concatenation can combine multi-view representations, it treats views equally and ignores inter-view dependencies. To overcome this limitation, we introduce a cross-view attention with consistency regularization that adaptively integrates complementary spatial cues from both projections to enhance robust VAT estimation. The code is available at: GitHub repository. The major contributions of this work are:

1. We present a multi-view, multi-modal framework for VAT estimation that integrates LS and PAS DXA scans with patient demographics via a cross-view attention module with consistency regularization to enable information sharing and aligned feature learning across projections.
2. We show that joint modelling of LS and PAS DXA scans captures complementary information relevant to VAT estimation, leading to improved performance compared to single-view approaches.
3. We further validate the clinical relevance of predicted VAT by showing strong associations with prevalent Atherosclerotic Cardiovascular Disease (ASCVD) and diabetes, supporting its use for opportunistic cardiometabolic risk assessment in the absence of direct VAT measurements.

2 Methodology

The proposed multi-view, multi-modal framework (see Figure 1) for estimating VAT from LS and PAS DXA scans and patient demographics comprises four main components: (a) dual-view image encoders, (b) a demographic data encoder, (c) a cross-view attention fusion module, and (d) a regression head.

Dual-view Image Encoders: The LS and PAS DXA scans provide complementary anatomical information for VAT estimation, so each view is processed by an independent convolutional encoder to preserve view-specific features.

Given a batch of LS images $X^{LS} \in \mathbb{R}^{B\times 3\times H_{LS}\times W_{LS}}$ and PAS images $X^{PAS} \in \mathbb{R}^{B\times 3\times H_{PAS}\times W_{PAS}}$, each view is encoded using an independent pretrained MobileNetv2 [13] backbone (truncated before the classifier). We avoid weight sharing between views to allow the encoders to learn view-specific representations that reflect differences in anatomical appearances across LS and PAS projections.

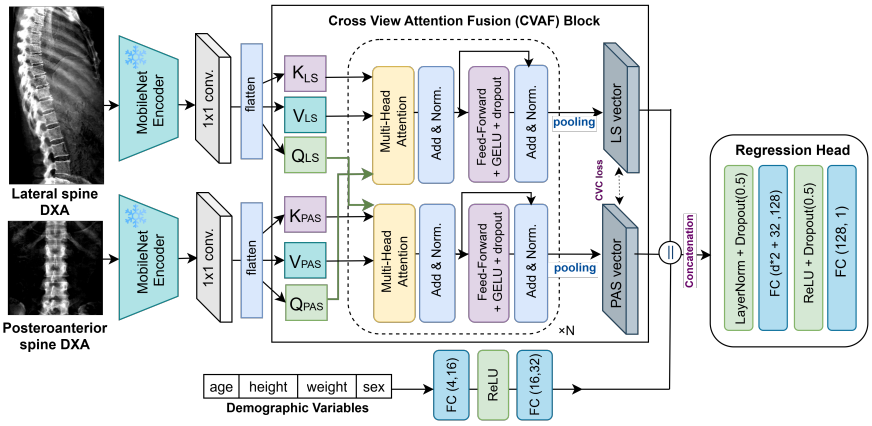

Fig. 1. Overview of the Proposed VisiFAT Framework.

Each encoder outputs a high-level feature map that preserves spatial structure while encoding rich semantic information.

Let ϕ_{LS} and ϕ_{PAS} denote the two feature encoders, respectively, initialized with ImageNet [4] weights and partially frozen (blocks 0–10) to stabilize training.

$$F^{\mathrm{LS}} = \phi_{\mathrm{LS}}\big(X^{\mathrm{LS}}\big) \in \mathbb{R}^{B\times C\times H_f\times W_f}; \quad F^{\mathrm{PAS}} = \phi_{\mathrm{PAS}}\big(X^{\mathrm{PAS}}\big) \in \mathbb{R}^{B\times C\times H_f\times W_f}.$$

Each encoder produces a feature tensor F^{LS}, F^{PAS} in $\mathbb{R}^{B\times C\times H_f\times W_f}$, where C=1280 is the number of output channels, and H_f, W_f denote the spatial dimensions of the final convolutional feature maps.

Demographic Data Encoder: While image features capture anatomical structure, VAT is also strongly influenced by patient-level factors such as age, sex, weight and height. To fuse this information, we introduce a demographic encoder that transforms standardized features into a compact embedding.

Height and weight enable the model to capture body-size related variation, while sex accounts for systematic differences in fat distribution across patient subgroups. Each standardized demographic feature is encoded by a 2-layer FC(4→16→32) with ReLU activations to obtain a 32-dim embedding.

Importantly, this demographic information is not fused early with image features; instead, it is introduced at a later stage, allowing the image branches to first focus on anatomical feature extraction and cross-view alignment.

Cross-view Attention Fusion Block: After extracting spatial feature maps from each image view, a key challenge is to integrate complementary information across projections while preserving spatial specificity. Conventional multi-view fusion strategies (i.e. concatenation, averaging, or late fusion) treat views as independent or equally informative, and therefore fail to model fine-grained, region-level correspondence between projections. Recent studies have demonstrated that cross mechanisms can effectively capture fine-grained inter-view/modality

relationships [17,19]. We introduce a Cross View Attention Fusion (CVAF) module that enables spatially-aware interaction between views, allowing each projection to selectively query anatomically relevant regions in the other view.

First, the feature maps F^{LS} and F^{PAS} are projected into a shared d-dimensional (d=128) embedding space using 1×1 convolution, enabling cross-view attention by aligning both views in a common embedding space while preserving spatial correspondence. The projected feature maps are then flattened into sequences of spatial tokens, where each token corresponds to a localized anatomical region.

The CVAF module performs bi-directional cross-attention (PAS $\leftrightarrow$ LS) between the two token sequences to explicitly model interactions between the LS and PAS views. Rather than fusing views through static operations (i.e. concatenation or averaging), cross-attention enables each view to dynamically query the other and selectively integrate information based on anatomical relevance. In each cross-attention layer, bi-directional cross-attention is performed. First, PAS tokens act as Queries (Q_{PAS}) while LS tokens serve as Keys (K_{LS}) and Values (V_{LS}). This allows each spatial region in the PAS view to identify and aggregate the most informative regions from the LS view, effectively answering the questions like: *which LS anatomical structures are most relevant for refining this PAS region?* Subsequently, LS tokens act as Q_{LS}, while PAS tokens provide the L_{PAS} and K_{PAS}, enabling reciprocal information exchange between views. To simplify, cross-attention from PAS to LS is computed as:

$$\mathrm{Attn}(Q_{\mathrm{PAS}}, K_{\mathrm{LS}}, V_{\mathrm{LS}}) = \mathrm{softmax}\left(\frac{Q_{\mathrm{PAS}} K_{\mathrm{LS}}^{\top}}{\sqrt{d}}\right) V_{\mathrm{LS}}, \tag{1}$$

Each cross-attention block applies layer normalization before attention, followed by a position-wise feed-forward network with GELU and dropout. A mean pooling layer then produces a fixed-length embedding per view. The resulting LS and PAS embeddings capture within- and cross-view interactions and are combined into a unified image representation for fusion with demographic embeddings.

Cross-View Consistency Regularization: Although the LS and PAS DXAs differ substantially in appearance and projection geometry, they capture the same underlying anatomy and therefore should encode consistent global semantic information relevant to VAT estimation. Without proper view-level alignment, the model can rely on view-specific cues that do not transfer well for generalization. To address this, we introduce a Cross-View Consistency (CVC) term during training that penalizes the squared distance between both views. Rather than enforcing identical embeddings or using contrastive objectives that explicitly separate views, the CVC loss softly encourages similarity while preserving view-specific information, leading to improved performance (see Table 2).

The overall training objective is defined as: $\mathcal{L} = \mathcal{L}_{\mathrm{SmoothL1}} + \lambda\, \mathcal{L}_{\mathrm{consistency}}$, where $\mathcal{L}_{\mathrm{SmoothL1}}$ denotes the primary regression loss on VAT mass and λ (here $\lambda = 0.01$) controls the contribution of the CVC term.

Regression Head: The fused image features are concatenated with a 32-dim demographic embedding and passed through a two-layer regression head (LayerNorm, dropout, ReLU) to produce the final VAT mass prediction.

3 Experiments and Results

Datasets: We used 37,772 de-identified and matched LS and PAS DXA scans (instance 2; 2014+) from UK Biobank[8] Imaging Study, a sub cohort of the large population-based cohort of approximately 500,000 participants (aged: 40 to 69 years). All the scans were captured using an iDXA bone density machine (GE Healthcare). The PAS scans had variable sizes ranging from (275 x 276) to (1036 x 720) pixels, while LS spine scans varied from (145 x 716) to (1962 x 808) pixels. All images were resized with preserved aspect ratio and zero-padded to a fixed input size to maintain anatomical consistency and minimize distortion. Demographic data (age, sex, weight, and height) was available for all participants. Visceral fat mass (0–8,631 grams) from whole-body DXA, quantified using the CoreScan software, served as the ground truth outcome, with CoreScan used as the reference standard due to its large-scale availability within UK Biobank. To the best of our knowledge, this is the only available dataset with paired LS and PAS DXA scans alongside visceral fat values. The data was accessed in accordance with UK Biobank data access regulations.

Implementation Details: All images from both LS and PAS views were resized to fixed dimensions of (600×277) and (384×384), respectively, while preserving their aspect ratio to avoid distorting anatomical structures. Single-channel DXA scans were replicated into three-channel inputs and normalized to match the input requirements of the pretrained backbone. Demographic features are z-score standardized (sex encoded 0/1). Training is conducted using stratified 10-fold cross-validation. Within each fold, the data is divided into 80% training, 10% validation, and 10% testing subsets, with strict separation to prevent leakage. The model is trained end-to-end for 30 epochs using AdamW with a learning rate of 3×10^{-4}, weight decay of 10^{-3}. To improve stability and generalization, we use an exponential moving average (decay 0.999) and a cosine learning-rate schedule with linear warmup. All experiments are implemented in PyTorch and run on two NVIDIA Quadro RTX 8000 GPUs. We evaluate model performance using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Symmetric Mean Absolute Percentage Error (sMAPE), and Pearson's Correlation.

Multi-view Baseline: Baseline employs two MobileNetV2 [13] encoders (one per view) with frozen early layers, concatenated with a 32-dim demographic data embedding and a regression head (FC(2592, 512), dropout(0.5), FC(512, 1)).

Results and Discussion: Table 1 presents a quantitative comparison of the proposed method, the single-view approach [10], and a multi-view baseline, using 10-fold averaged metrics on the UK Biobank dataset under the same training protocol. Compared to multi-view baseline, the proposed model achieves a 21.0% reduction in RMSE, a 21.7% reduction in MAE, and a 30.9% reduction in sMAPE. Relative to the single-view approach, the corresponding reductions are 31.5% (RMSE), 33.5% (MAE), and 38.1% (sMAPE). These results demonstrate substantial gains across all evaluation metrics.

[8] https://www.ukbiobank.ac.uk/

Table 1. Comparison with the single-view approach [10] and multi-view baseline.

Model	Input Configuration			RMSE (g) ↓	MAE (g) ↓	sMAPE ↓
	PAS	LS	demographic data			
Maqsood et al. [10]	×	✓	age, weight, height	327.64 ±6.65	231.29 ±7.82	28.86% ±1.36
Multi-view Baseline	✓	✓	**sex**, age, weight, height	284.14 ±4.86	196.44 ±2.86	25.83% ±0.88
Proposed	✓	✓	**sex**, age, weight, height	224.41±3.89	153.82 ±3.23	17.85% ±0.43

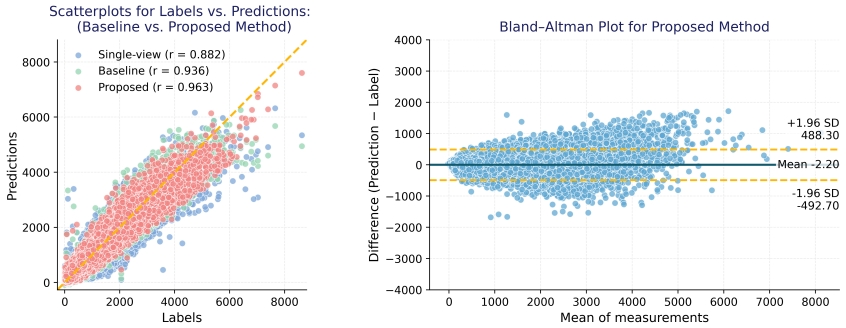

Fig. 2. Comparison scatterplots for UK Biobank using the single-view approach, baseline, and proposed model, along with a Bland–Altman plot for proposed model.

The scatterplot (Figure 2, left) shows strong agreement between predicted and true values, with the highest Pearson correlation (0.963) for the proposed model. The Bland–Altman plot (Figure 2, right) indicates a near-zero mean bias, indicating minimal systematic error across the measurement range, with most predictions falling within the 95% limits of agreement. These results demonstrate that naively incorporating multiple views is insufficient to fully exploit the complementary information present in LS and PAS DXA images. While the multi-view baseline leverages both projections, it lacks explicit inter-view modelling. The proposed framework enhances VAT estimation by enabling rich spatial feature exchange through cross-view attention.

Ablation Studies: We compare input configurations and multi-view fusion under identical settings for fair evaluation. Table 2 compares different feature fusion strategies for both DXA views. The best performance is achieved when cross-view attention is combined with CVC regularization, confirming the benefit of modelling cross-view interactions and enforcing consistency across views.

Table 2. Comparison of fusion techniques for combining multi-view image features.

Fusion Technique	RMSE (g) ↓	MAE (g) ↓	sMAPE
Concatenation (multi-view baseline)	284.14	196.44	25.83%
Gated Fusion	267.99	189.31	23.78%
Cross-view Attn.	249.42	178.08	21.85%
Cross-view Attn. + CVC regularization	224.41	153.82	17.85%

Table 3 demonstrates the impact of different input configurations on model performance. Combining LS and PAS views with patient demographic data achieves the best performance, highlighting the complementary nature of multi-view imaging and auxiliary patient demographic information.

Table 3. Evaluation of Different Input Configurations.

Input Configuration	**RMSE** (g) ↓	**MAE** (g) ↓	**sMAPE** ↓
demographic data (sex, age, weight, height)	505.47	382.09	46.58%
LS only	353.21	275.82	29.73%
LS + demographic data	327.64	231.29	28.86%
PAS only	359.97	291.32	30.31%
PAS + demographic data	321.98	236.69	29.58%
LS + PAS	276.33	198.11	20.42%
LS + PAS + demographic data	224.41	153.82	17.85%

Clinical Analysis: Given the strong association between VAT and cardio-metabolic health [14], we examined the cross-sectional relationships of both ground truth and estimated VAT with prevalence of ASCVD and diabetes at the time of imaging as separate outcomes. Multi-variable-adjusted logistic regression models were applied in individuals with available VAT predictions and corresponding ground truth measurements (n=35,330, 50.7% female, mean age 64.5 ± 7.9 years). Notably, there were an additional 12,963 individuals (51.3% female, mean age 65.0 ± 7.6 years) with only VAT predictions and complete co-variate data (termed 'unseen', with no ground truth VAT available). The date of DXA imaging was considered the index date, with any ASCVD or diabetes event prior to this considered prevalent. Both ground truth and estimated VAT was also categorized based on sex-specific tertiles (lowest, moderate and highest). Prevalent ASCVD was defined as a composite outcome comprising of myocardial infarction, stroke, or coronary artery disease. Prevalent diabetes was also obtained from linked health records. Detailed definitions of outcomes and

Table 4. Odds ratios for the multi-variable-adjusted relationship between estimated VAT tertiles, prevalent ASCVD and diabetes.

	Cohort			**Tertile 1**	**Tertile 2**	**Tertile 3**
Prevalent ASCVD	**Training** (n=35,330)	Predicted VAT	Prevalence	323 (2.7)	453 (3.9)	631 (5.4)
			MV-model	Ref.	1.24 (1.07-1.44)	1.55 (1.35-1.80)
		Ground truth VAT	Prevalence	312 (2.7)	448 (3.8)	467 (4.0)
			MV-model	Ref.	1.25 (1.07-1.45)	1.65 (1.43-1.91)
	Unseen (n=12,963)	Predicted VAT	Prevalence	136 (3.2)	213 (4.9)	616 (4.8)
			MV-model	Ref.	1.43 (1.14-1.79)	1.66 (1.33-2.07)
Prevalent Diabetes	**Training** (n=35,330)	Predicted VAT	Prevalence	121 (1.0)	250 (2.1)	757 (6.4)
			MV-model	Ref.	1.84 (1.48-2.29)	5.51 (4.51-6.72)
		Ground truth VAT	Prevalence	112 (1.0)	258 (2.2)	758 (6.4)
			MV-model	Ref.	2.05 (1.64-2.57)	5.99 (4.88-7.36)
	Unseen (n=12,963)	Predicted VAT	Prevalence	58 (1.3)	104 (2.4)	276 (6.4)
			MV-model	Ref.	1.61 (1.16-2.24)	4.23 (3.15-5.69)

MV-model adjusted for age, sex, ethnicity, systolic blood pressure, physical activity, smoking, country of residence, scan year, and prevalent ASCVD or diabetes (depending on outcome).

co-variates are provided in [16]. Scan year was used as a variable to account for potential temporal acquisition variability. Table 4 shows that observed associations between VAT and cardio-metabolic outcomes remain consistent across

analyses, with higher VAT levels associated with progressively greater odds of prevalent ASCVD and diabetes.

4 Conclusion

We propose a multi-view, multi-modal deep learning framework for VAT estimation from DXA scans that leverages complementary LS and PAS projections with demographic data. The model integrates cross-view attention for dynamic feature exchange and a consistency regularization strategy to align latent representations, improving robustness and generalization. Extensive experiments show consistent gains over single-view and baseline fusion methods. The estimated VAT also demonstrates strong, significant associations with ASCVD and diabetes, highlighting both methodological and clinical relevance.

Data Use Declaration: Ethical Approval (Application ID 93712) for the UK Biobank was obtained from the Northwest Multi-center Research Ethics Committee (United Kingdom, Ref: 21/NW/0157). Written informed consent was also obtained from each participant (Ref: 11/NW/03/820). The study was supported by a National Health and Medical Research Council (NHMRC) of Australia Ideas grant (APP1183570) and a Medical Research Future Fund 2022 Cardiovascular Health Mission grant (MRF2024225). M.S was supported by a project grant from the Western Australian Future Health Research and Innovation Fund. J.R.L was supported by National Heart Foundation of Australia Future Leader Fellowship (102817 & 107323). S.Z.G. received partial support from the WA Department of Health Near Miss Award (G1008050). N.C.H. was supported by the UK Medical Research Council [MC_PC_21003 & 21001] and the NIHR Southampton Biomedical Research Centre, UK.
Disclosure of Interests: The authors have no competing interests to declare that are relevant to content of this paper.

References

1. Ahmed, S.K., Mohammed, R.A.: Obesity: Prevalence, Causes, Consequences, Management, Preventive Strategies and Future Research Directions. Metabolism Open p. 100375 (2025)
2. Butler, J., Fioretti, F., Davies, M.J.: Redefining Obesity: Beyond Body Mass Index (2025)
3. 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
4. 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)
5. Endocrinology, T.L.D.: Redefining Obesity: Advancing Care for Better Lives (2025)
6. Huffman, D.M., Barzilai, N.: Role of Visceral Adipose Tissue in Aging. Biochimica et Biophysica Acta (BBA)-General Subjects **1790**(10), 1117–1123 (2009)

7. 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
8. Karastergiou, K., Smith, S.R., Greenberg, A.S., Fried, S.K.: Sex Differences in Human Adipose Tissues–The Biology of Pear Shape. Biology of Sex Differences **3**(1), 13 (2012)
9. Kim, H., Kim, S.E., Sung, M.K.: Sex and Gender Differences in Obesity: Biological, Sociocultural, and Clinical Perspectives. The World Journal of Men's Health **43**(4), 758 (2025)
10. 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., et al.: From Pixels to Prognosis: A Multi-modal Attention-Based Framework for Visceral Adipose Tissue Estimation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 249–259. Springer (2025)
11. 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
12. 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
13. 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)
14. Shetty, S., Suvarna, R., Bhattacharya, S., Seetharaman, K.: Visceral adiposity and cardiometabolic risk: clinical insights and assessment. Cardiology in Review pp. 10–1097 (2025)
15. 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), 1–10 (2012)
16. Sim, M., Webster, J., Smith, C., Saleem, A., Gilani, S.Z., Toro-Huamanchumo, C.J., Suter, D., Figtree, G., Lagendijk, A.K., Duncan, E.L., Schultz, C., Szulc, P., Hung, J., Lim, W.H., Raina, P., Bondonno, N.P., Woodman, R., Hodgson, J.M., Kiel, D.P., Prince, R.L., Lewis, J.R.: Automated Abdominal Aortic Calcification Scores and Atherosclerotic Cardiovascular Disease in the UK Biobank Imaging Study. JACC: Advances **5**(3), 102570 (2026). https://doi.org/10.1016/j.jacadv.2025.102570
17. Tariq, A., Maqsood, A., Masek, M., Gilani, S.Z.: BiFuseNet: A Multimodal Network for Estimating Blood Alcohol Concentration via Bidirectional Hierarchical Fusion. In: Proceedings of the 27th International Conference on Multimodal Interaction. pp. 605–613 (2025)
18. World Health Organization: Obesity and Overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (2025)
19. Zhou, B., Krähenbühl, P.: Cross-view Transformers for Real-time Map-view Semantic Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13760–13769 (2022)