Journal of Imaging Informatics in Medicine
https://doi.org/10.1007/s10278-026-02046-3

Dynamic Fuzzy-Gaussian Modeling (DynFGM): A Kurtosis-Adaptive Unsupervised Framework for Automated Adipose Tissue Segmentation in Abdominal MRI

Asefa Adimasu Taddese[1] · Joshua D. K. Bernal[2] · Chit K. Leung[2] · Yu-Ying Wang[1] · Angus P. Yu[3] · Zuneera Umair[4] · Bjorn T. Tam[1,5]

Received: 11 April 2026 / Revised: 23 May 2026 / Accepted: 1 June 2026

Abstract

Accurate MRI-based quantification of abdominal adipose tissue is critical for metabolic risk assessment but is limited by labor-intensive manual segmentation and the extensive labeled-data dependency of deep learning models. We introduce Dynamic Fuzzy-Gaussian Modeling (DynFGM), a fully automated, unsupervised framework for adipose tissue segmentation designed to operate without requiring training data, expert annotations, or anatomical priors. DynFGM was developed and validated on 776 abdominal MRI scans, using a benchmark cohort ($n = 20$) with expert ground truth segmentations and a large validation cohort ($n = 756$). The pipeline dynamically adapts its complexity for each MRI slice by using image intensity kurtosis to select the optimal number of tissue clusters. A fuzzy *C*-means (FCM) algorithm then initializes a Gaussian mixture model (GMM) for segmentation, providing a mathematically interpretable alternative to black-box neural networks. Finally, a radial distance transform with an adaptive cutoff differentiates subcutaneous (SAT) from visceral adipose tissue (VAT). Performance was evaluated against the ground truth using dice similarity coefficient (DSC) and intraclass correlation coefficient (ICC). DynFGM achieved strong spatial agreement with expert annotations (mean DSC: 0.94) and high volumetric reliability (ICC: 0.82–0.97), comparable to reported inter-expert variability. The framework reduced mean absolute volumetric error by 92.6% compared to standard FCM (482.2 cm^3 vs. 6547.5 cm^3). On the large validation cohort ($n = 756$), the method demonstrated operational stability, producing physiologically plausible adipose distributions with a low technical failure rate (3.0%). Furthermore, the computational throughput averaged 13.6 s per participant on standard CPU (Intel® Core™ i9, 3.0 GHz) hardware. DynFGM provides an interpretable and data-efficient approach for abdominal adipose tissue phenotyping, offering an alternative to supervised deep learning in settings where labeled data are limited or unavailable. By bridging the gap between manual segmentation and labeled-data-dependent AI, this unsupervised framework offers a scalable tool for population-level research and may serve as an automated labeling tool to facilitate future model development.

Keywords Dynamic Fuzzy-Gaussian Modeling · Abdominal adipose tissue · Kurtosis adaptation · MRI segmentation · Imaging informatics · Unsupervised learning

✉ Bjorn T. Tam
bjorntam@hkbu.edu.hk

[1] Academy of Wellness and Human Development, Faculty of Arts and Social Sciences, Hong Kong Baptist University, Hong Kong SAR, China

[2] Division of Kinesiology, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

[3] Department of Sports Science and Physical Education, The Chinese University of Hong Kong, Hong Kong SAR, China

[4] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China

[5] Dr. Stephen Hui Research Centre for Physical Recreation and Wellness, Faculty of Arts and Social Sciences, Hong Kong Baptist University, Hong Kong SAR, China

Published online: 22 June 2026

Introduction

Excess visceral adipose tissue (VAT) is a primary driver of metabolic disease and cardiovascular risk [1, 2]. Elevated VAT volume and visceral to subcutaneous fat ratios have a substantial correlation with insulin resistance, dyslipidemia, and systemic inflammation [3–5]. Magnetic resonance imaging (MRI) is widely recognized as the gold standard for the non-invasive quantification of these depots, providing high-contrast, three-dimensional biomarkers of obesity-related health [4, 6, 7]. However, routine extraction of volumetric data from abdominal MRI is frequently bottlenecked by the manual segmentation process, which is labor-intensive and prone to observer variability [8]. These limitations hinder the integration of MRI-based adipose phenotyping into large-scale clinical and epidemiological studies, underscoring the need for accurate, automated, and reproducible segmentation frameworks that can operate robustly across heterogeneous datasets.

A variety of automated approaches have been developed to address this challenge, each offering advantages in specific research contexts. Traditional threshold-based and atlas-guided methods provide computational efficiency and can provide anatomically consistent segmentations in standardized settings. Nevertheless, their performance often degrades in the presence of protocol variation, intensity inhomogeneity, or atypical body habitus [7, 9]. Unsupervised clustering techniques, such as fuzzy *C*-means (FCM), represented an important advance by enabling label-free tissue separation. However, these methods typically rely on a fixed number of tissue clusters (K), and this inherent rigidity can lead to segmentation inaccuracies when image contrast or anatomical composition varies across slices or participant cohorts [10, 11]. Spatially constrained and hybrid FCM variants have improved robustness in some scenarios [8, 10, 12], but they often require extensive parameter tuning and remain sensitive to scanner- and site-specific differences, limiting widespread deployment in multi-center studies.

In recent years, deep learning (DL) architectures, such as U-Net and nnU-Net, have achieved state-of-the-art segmentation performance, often reaching human-level accuracy [13, 14]. These models perform exceptionally well in data-rich environments where large, expertly annotated training datasets are available [15]. However, their broader clinical adoption is constrained by the high cost of generating pixel-level annotations, their dependence on large training corpora, and their limited interpretability compared with transparent statistical methods [2, 6, 16]. Consequently, there remains a clear need for segmentation methods that are accurate, fully automated, interpretable, and capable of operating without labeled training data.

Abdominal fat–water MRI further amplifies these technical challenges, as images are frequently affected by bias field artifacts, partial-volume effects, and intrinsic biological heterogeneity that induce slice-to-slice intensity fluctuations [2, 17]. Fixed-complexity segmentation strategies, whether conventional FCM or pre-trained neural networks, may fail to adapt to these variations, particularly in anatomically complex regions such as the diaphragm, abdominal wall, and pelvic inlet [18, 19]. This issue is especially problematic in large-scale or multi-site studies, where acquisition parameters, body shapes, and depot morphology can vary substantially. A key methodological gap, therefore, lies between conventional unsupervised clustering methods, which lack the flexibility to adapt segmentation complexity to local anatomy, and supervised DL approaches, which require extensive annotated data and often lack transparency.

To address these limitations, the present study introduces a Dynamic Fuzzy-Gaussian Modeling (DynFGM), a fully automated and interpretable unsupervised framework for abdominal MRI segmentation that dynamically adapts segmentation complexity according to slice-wise image characteristics. The core innovation of the method is a kurtosis-adaptive mechanism that adjusts the number of tissue clusters on a per-slice basis. By using image intensity kurtosis, a statistical measure of distribution tailedness, as a heuristic for local anatomical complexity, DynFGM automatically selects the appropriate segmentation granularity for each slice. Although kurtosis is well established in diffusion-weighted MRI and other quantitative imaging contexts [20, 21], its use here to drive dynamic clustering complexity provides a novel strategy for tailoring unsupervised segmentation to heterogeneous anatomy.

By integrating fuzzy *C*-means (FCM) with probabilistic Gaussian mixture refinement, the proposed framework combines uncertainty-aware boundary modeling with robust statistical classification while avoiding dependence on labeled training data or atlas-based anatomical priors. FCM is employed to generate soft tissue memberships that capture boundary uncertainty and provide a robust initialization of cluster means and covariances. These parameters are then refined using an Expectation–Maximization (EM) algorithm to estimate a GMM that is resilient to noise and intensity inhomogeneity, while preserving mathematical interpretability. Finally, a radial distance transform with an adaptive cutoff is applied to separate SAT from VAT, leveraging geometric relationships to derive anatomically plausible compartment labels without reliance on atlas priors or training data. The main contributions of this study are threefold. First, it introduces a kurtosis-driven, dynamically adaptive clustering framework for abdominal MRI that operates fully unsupervised, eliminating the need for expert annotations, atlas registration, or site-specific calibration.

Second, it proposes a hybrid FCM–GMM segmentation pipeline that explicitly models uncertainty and intensity distributions, providing a transparent alternative to black-box neural networks while achieving high agreement with expert segmentations. Third, it demonstrates, in a large MRI cohort, that DynFGM produces physiologically plausible segmentation and quantification with low failure rates and clinically feasible processing times, thereby offering a practical solution for population-scale adipose phenotyping.

By bridging the gap between rigid conventional clustering and data-hungry deep learning, DynFGM provides a statistically interpretable and computationally efficient tool that can function both as a standalone segmentation framework for novel datasets and as an automated labeling engine to support future DL model development.

Methods

Study Design and Cohorts

This methodological study was validated using 776 participant scans, with participants aged 20–72 years who met Asian-specific criteria for overweight (body mass index ≥ 23) and central obesity (waist circumference ≥ 90 cm for men and ≥ 80 cm for women) [22]. All participants underwent abdominal MRI for body composition assessment between 2021 and 2024. The MRI datasets analyzed herein were derived as secondary data from a previously conducted clinical trial (ClinicalTrials.gov ID: NCT04887454) [23]. All data were fully anonymized prior to analysis. In accordance with institutional and local regulatory guidelines, the reuse of anonymized secondary data for methodological development did not require additional ethical approval.

We employed a two-tier validation design. (1) Benchmark cohort ($n = 20$): A randomly selected subset of participants was used for quantitative accuracy assessment. For these cases, VAT and SAT were manually segmented on 24 axial slices centered at the L3–L5 vertebral levels—an established anatomical landmark for abdominal adiposity quantification [24]. Segmentations were independently performed by two experienced researchers using 3D Slicer software [25], under the supervision of a certified radiologist. Discrepancies were resolved through consensus, and the resulting annotations served as the ground truth reference. (2) Validation cohort ($n = 756$): The remaining participants, each contributing 24 axial slices, were used to evaluate the operational stability, result plausibility, and processing throughput of the unsupervised framework across a large and heterogeneous population.

MRI Acquisition and Reconstruction

All abdominal MRI scans were acquired using a 3.0 Tesla GE Discovery 750 W system (GE HealthCare, USA), with the IDEAL IQ sequence (Iterative Decomposition of water and fat with Echo Asymmetry and Least squares estimation) implemented as a gradient-echo fat–water separation method optimized for adipose-tissue quantification. The imaging protocol included a repetition time (TR) of 160 ms (ms), dual echo times ($TE_1 = 2.1$ ms, $TE_2 = 3.15$ ms), flip angle of 85°, and spectral fat suppression using Spectral Adiabatic Inversion Recovery (SPAIR). Axial slices of 10-mm (mm) thickness were acquired in an interleaved manner, with an in-plane field of view of 420×380 mm^2, matrix size 256×160, and voxel size of $1.6 \times 2.4 \times 10$ mm^3. The reconstruction yielded four co-registered volumes: water-only, fat-only, in-phase, and out-of-phase images. Participants fasted for ≥ 4 h prior to scanning to minimize metabolic variability in the fat signal.

Image Preprocessing

DICOM images were converted to 16-bit grayscale arrays and preprocessed slice-by-slice. The Rudin–Osher–Fatemi (ROF) total variation (TV) denoising model [26] was applied to suppress Rician noise while preserving anatomical boundaries, minimizing the energy function, with a regularization weight of $\lambda = 0.01$ (empirically chosen for optimal denoising).

$$\min_u \lambda \int \|\nabla u\| dx + \frac{1}{2}\|u - I_{noisy}\|_2^2$$

where u denotes the denoised image, I_{noisy} represents the observed noisy image, ∇u is the image gradient, and λ controls the balance between edge preservation and noise suppression.

Subsequently, intensity nonuniformity was corrected using N4 bias field correction [27]. Finally, Contrast-Limited Adaptive Histogram Equalization (CLAHE; clip-limit = 2.0, tile grid size = 8×8) was applied, ensuring only approximately 1% of pixel values were saturated, thereby enhancing local tissue contrast for subsequent clustering [28].

Abdominal Body Mask Generation

To exclude background regions and focus analysis on the anatomical volume, a binary body mask was computed for each axial slice. Otsu's thresholding was selected for its robustness and O(N) histogram-based computation, providing a statistically optimal foreground–background separation

that is computationally efficient for high-throughput, full-volume MRI analysis [29].

The resulting mask was refined using morphological closing with a disk-shaped structuring element (radius = 2 pixels) to bridge narrow gaps, followed by hole filling to generate a continuous body region. To retain only the largest contiguous object and exclude air-filled regions and peripheral artifacts, connected-component labeling was applied.

Dynamic Fuzzy-Gaussian Modeling (DynFGM)

The DynFGM framework integrates slice-wise kurtosis-guided cluster selection, fuzzy *C*-means (FCM) initialization, and Expectation–Maximization (EM)-based Gaussian mixture model (GMM) refinement. This method leverages fuzzy logic to handle voxel assignment uncertainty and probabilistic modeling for statistical robustness, addressing key challenges in MRI segmentation such as noise, intensity heterogeneity, and partial-volume effects, particularly relevant for AI-driven biomedical image analysis.

Automated Cluster-Number Selection

For each body-masked slice, the Pearson kurtosis (κ) of the voxel intensity distribution was computed as a measure of distributional tailedness:

$$\kappa = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(I_i - \overset{\}}{I}\right)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(I_i - \overset{\}}{I}\right)^2\right)}$$

where I_i is the intensity of voxel i, $\overset{\}}{I}$ is the mean intensity, and N is the total number of voxels in the body mask. We utilized excess kurtosis ($\kappa - 3$) as a quantitative proxy for anatomical complexity (tissue heterogeneity and image contrast), where a positive value indicates a leptokurtic (peaked) distribution and a negative value indicates a platykurtic (flat) distribution.

The number of tissue clusters (K) for the segmentation model was then dynamically assigned using the following rule:

$$K = \begin{cases} 2, \ if\ \kappa - 3 > 0.5 \ (strongly\ leptokurtic\ (homogeneous) \\ 3, \ if\ -0.5 \le \kappa - 3 \le 0.5 \ (mesokurtic\ (normal\ complexity) \\ 4, \ if\ \kappa - 3 < -0.5\ strongly\ platykurtic\ (heterogeneous) \end{cases}$$

This rule encodes the adaptive logic of the proposed framework. The thresholds of −0.5 and +0.5 were selected because they effectively partitioned the kurtosis distribution into three distinct modes, corresponding to slices with low (homogeneous, e.g., mid-abdomen), intermediate (e.g., involving multiple organ boundaries), and high (e.g., at the diaphragm or iliac crest) anatomical complexity.

Slices with high positive excess kurtosis (> 0.5) exhibited simple, bimodal intensity distributions, for which a two-cluster model ($K = 2$; fat and non-fat) was sufficient. Near-zero excess kurtosis ($-0.5 \le \kappa - 3 \le 0.5$) indicates a more normally distributed yet complex intensity profile, warranting a three-cluster model ($K = 3$) to potentially separate SAT, VAT, and background/muscle. Conversely, low negative excess kurtosis (< -0.5) signifies a flat, highly heterogeneous distribution with overlapping tissue intensities, commonly observed at the diaphragm or iliac crest. For these challenging slices, a four-cluster model ($K = 4$) provides the granularity needed to resolve SAT, VAT, muscle, and background/organs.

This adaptive approach matches model complexity to the inherent data complexity on a slice-by-slice basis. The stability of the final segmentation results with respect to these threshold choices is demonstrated in our sensitivity analysis.

FCM-Initialized GMM Clustering

Phase A: FCM Initialization FCM clustering was applied with the selected K, minimizing the objective function:

$$J = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}^{m} \parallel x_i - c_k \parallel^2$$

where u_{ik} is the membership degree of voxel i in cluster k, x_i denotes the intensity value of voxel i, c_k is the cluster center, and $m = 2$ is the fuzzifier.

Iterative updates proceeded until convergence:

$$u_{ik} = \left(\sum_{j=1}^{K}\left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|}\right)^{2/(m-1)}\right)^{-1} \quad c_k = \frac{\sum_{i=1}^{N} u_{ik}^{m} x_i}{\sum_{i=1}^{N} u_{ik}^{m}}$$

This step manages boundary uncertainty and partial-volume effects, providing a robust starting point for probabilistic refinement.

Phase B: GMM Refinement To refine the membership assignments generated by FCM, a one-dimensional GMM was fitted to the voxel intensities within the body mask. GMMs provide a statistically rigorous framework for MRI segmentation by modeling intensity heterogeneity as a linear combination of Gaussian distributions [30]. The initial parameters for the GMM, means (μ_k), variances (σ_k^2), and mixing proportions (π_k), were seeded using the cluster centers (c_k) derived from the FCM stage, with $\pi_k = 1/K$ for all k.

The intensity distribution of the image was modeled as follows:

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(x|\mu_k, \sigma_k^2\right) \ \text{subjectto} \sum \pi_k = 1$$

where $p(x)$ represents the probability density function of voxel intensity x, π_k are the mixture weights satisfying $\sum_{k=1}^{k}\pi_k=1$, and $N(\cdot)$ denotes the gaussian distribution for the k^{th} component.

Parameter estimation was performed using the Expectation–Maximization (EM) algorithm:

E-step: Compute posterior probabilities (w_{ik}) of voxel i belonging to component k

$$w_{\rangle k} = \frac{\pi_k . \mathcal{N}\left(x_i|\mu_k, \sigma_k^2\right)}{\sum_{|=1}^{K}\pi_{|} . \mathcal{N}\left(x_i|\mu_{|}, \sigma_{|}^2\right)}$$

M-step: Update parameters using posterior-weighted statistics

$$\mu_k = \frac{\sum_i w_{ik} I_i}{\sum_i w_{ik}},\ \pi_k = \frac{1}{N}\sum_i w_{ik},\ \sigma_k^2 = \frac{\sum_i w_{ik}\left(I_i-\mu_k\right)^2}{\sum_i w_{ik}}$$

Iterations proceeded until the log-likelihood convergence threshold (10^{-6}) was reached. Each voxel was then assigned to the component with the maximum posterior probability, yielding a refined hard classification map.

Fat-Mask Extraction and Morphological Refinement

Following GMM classification, the component representing adipose tissue was identified by selecting the cluster with the second-highest mean intensity (to account for potential high-intensity artifacts or noise in the brightest cluster). All other clusters were labeled as non-adipose (muscle, organs, or background). The resulting binary mask was refined through a series of morphological operations: (1) hole filling to account for intra-tissue intensity variations; (2) connected-component labeling to retain only the largest contiguous object, effectively removing spurious speckles (<2 px^2); and (3) closing and dilation (1–2 voxel radius) to recover boundary voxels potentially lost during thresholding.

SAT/VAT Separation and Volume Estimation

To differentiate SAT from VAT without anatomical priors, we utilized a radial distance transform $D(x, y)$. First, a distance map was computed from the outer boundary of the convex-hull-filled body mask. Then, to account for inter-individual variability, the separation depth was not fixed but was determined by an adaptive percentile (P) calculated for each slice:

$$P = clamp\left(35 + \frac{\left|M_{fat}\right|}{2000}, 45{,}80\right)$$

Here, $|M|_{fat}$ is the total number of fat voxels in the slice and *clamp* (…) constrains the output value within the specified lower and upper bounds. The final separation threshold ($\lceil_{th}$) was set to the P-th percentile of the distance values for all fat voxels and defined as follows:

$$\lceil_{th} = percentile\left(\left\{D_{(p)}|p\epsilon M_{fat}\right\}, P\right)$$

Voxels where $D(x, y) < d_{th}$ were classified as SAT, while those where $D(x, y) \geq d_{th}$ were classified as VAT. A fixed-percentile cutoff is unreliable across diverse body compositions. This formula was therefore developed as an empirical model to approximate the anatomical location of the abdominal wall fascia.

Volumes were calculated using DICOM metadata for pixel spacing (dx, dy) and slice thickness ($dz(inmm)$). The voxel volume (v_{vox} = dx.dy.dz) was multiplied by the total voxel count for each compartment:

$$V_{SAT} = v_{vox}\sum_x M_{SAT}(x), V_{VAT} = v_{vox}\sum_x M_{VAT}(x)$$

Comparative Baseline Models

To isolate the performance gains of kurtosis-adaptive clustering and hybrid initialization, two conventional unsupervised models were implemented using the same preprocessing and body-masking pipeline:

1. Fixed-cluster FCM: applied independently to each slice with $k=4$ clusters, using a standard Euclidean distance metric and a fuzziness exponent $m=2$.
2. Standard GMM: utilized four Gaussian components initialized via k-means and trained using the Expectation–Maximization (EM) algorithm until log-likelihood convergence.

The complete methodology including image acquisition, image preprocessing, unsupervised model execution, and final quantitative segmentation outputs is schematically outlined in Fig. 1.

Evaluation Metrics and Statistical Analysis

Performance was rigorously evaluated on the benchmark cohort by comparing DynFGM and baseline outputs against expert-consensus segmentations (ground truth). Quantitative agreement was assessed using multiple complementary metrics: (1) the dice similarity coefficient (DSC) quantified voxel-wise segmentation accuracy; (2) mean absolute error (MAE) and mean squared error (MSE) characterized absolute and squared volume deviations across adipose compartments. (3) Spearman's rank correlation coefficient (ρ) assessed monotonic agreement, and intraclass correlation coefficients ($ICC_{2,1}$) using a two-way random-effects model

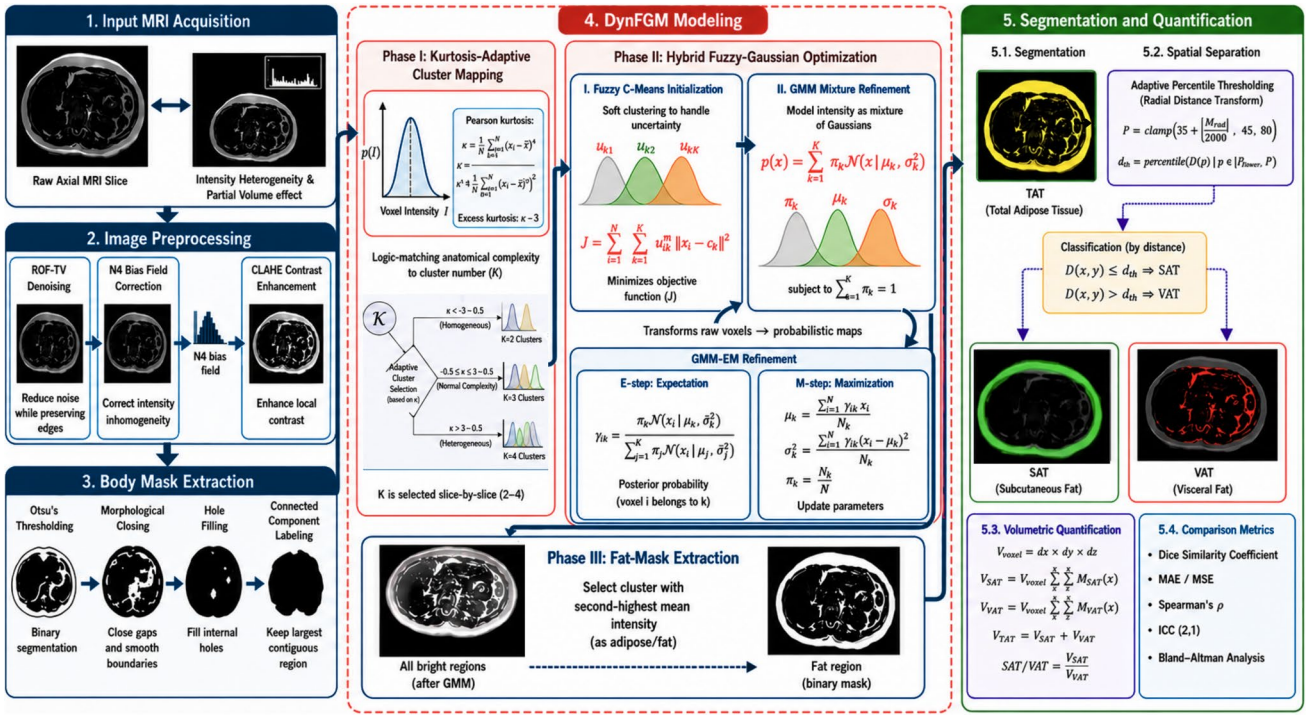

Fig. 1 Workflow of the proposed Dynamic Fuzzy-Gaussian Modeling (DynFGM) framework for automated abdominal adipose tissue segmentation and quantification in MRI. The workflow consists of five major stages: (1) MRI acquisition, where raw axial MRI slices containing intensity heterogeneity and partial volume effects are obtained; (2) image preprocessing, including ROF-TV denoising, N4 bias field correction, and CLAHE-based contrast enhancement; (3) body mask extraction using Otsu thresholding, morphological closing, hole filling, and connected-component labeling; (4) DynFGM modeling, comprising three phases: kurtosis-adaptive cluster mapping for adaptive cluster selection, hybrid fuzzy–Gaussian optimization using fuzzy *C*-means (FCM) and Gaussian mixture model–expectation maximization (GMM-EM) refinement, and adipose/fat mask extraction; and (5) segmentation and quantitative analysis of total adipose tissue (TAT), subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT), followed by volumetric quantification and statistical comparison metrics. Spatial separation between SAT and VAT is achieved using adaptive percentile thresholding based on radial distance transformation

assessed absolute agreement. (4) Bland–Altman analysis was used to estimate systematic bias and 95% limits of agreement (LoA).

Statistical significance was defined a priori as $p < 0.05$. Correlation strength was interpreted as follows: $|\rho| \geq 0.70$ (strong), 0.40–0.69 (moderate), and 0.10–0.39 (weak). In the validation cohort ($n = 756$), clinical plausibility was assessed by examining volume distributions and SAT: VAT ratios against physiologically plausible ranges.

Computational Performance and Reproducibility

All experiments were conducted on a standard CPU (Intel® Core™ i9, 3.0 GHz) with 64 GB DDR5 RAM (Windows 11, v24H2). Runtime was measured from image ingestion to final segmentation output.

The DynFGM framework and baseline models were implemented in Python 3.11 utilizing the following libraries: NumPy, SciPy, scikit-learn, scikit-image, SimpleITK, and fcmeans.

Results

Volumetric Accuracy and Spatial Agreement

In the benchmark cohort ($n = 20$), the DynFGM framework demonstrated high spatial agreement with expert consensus segmentations for total adipose tissue (TAT), achieving a DSC of 0.94 ± 0.03. DynFGM substantially outperformed conventional unsupervised methods in volumetric accuracy across all adipose compartments (Table 1). For TAT, DynFGM achieved a MAE of 482.2 cm^3, representing a 92.6% representing a 92.6% and 87.2% reduction compared to FCM and a standard GMM, respectively. For the clinically critical VAT depot, DynFGM achieved a MAE of 539.7 cm^3, significantly lower than both FCM (MAE = 3509.9 cm^3; $p < 0.01$) and GMM (MAE = 1127.3 cm^3; $p < 0.01$). While the error reduction in the SAT was also significant compared to baselines, the absolute MAE remained higher (778.4 cm^3) than that of

Table 1 Quantitative performance comparison of adipose tissue segmentation methods

Tissue	Method	MAE (cm^3)	MSE (cm^6)
TAT	FCM	6547.45	4.48×10^7
	GMM	3758.94	1.49×10^7
	DynFGM	482.24[a]	3.97×10^5
SAT	FCM	3036.70	9.44×10^6
	GMM	2630.81	7.39×10^6
	DynFGM	778.42[a]	8.17×10^5
VAT	FCM	3509.95	1.35×10^7
	GMM	1127.33	1.96×10^6
	DynFGM	539.73[a]	4.59×10^5

Volumetric accuracy of the proposed DynFGM against baseline FCM and GMM models across three primary adipose compartments. Performance is quantified using mean absolute error (MAE) and mean squared error (MSE), representing the average volumetric deviation from expert-consensus ground truth

[a]Significant reduction in error compared to baseline FCM and GMM models ($p < 0.05$). The DynFGM framework achieves a 92.6% reduction in volumetric error compared to standard clustering techniques for total adipose tissue

the visceral depot, reflecting the increased complexity of subcutaneous boundary delineation.

Agreement and Reliability

DynFGM volume estimates demonstrated strong monotonic agreement with expert-defined rankings. Spearman's correlation (ρ) was 0.99 for TAT, outperforming FCM ($\rho = 0.70$) and GMM ($\rho = 0.79$). For VAT, DynFGM achieved $\rho = 0.90$, compared to 0.41 (FCM) and 0.76 (GMM). These results confirmed that DynFGM reliably preserves volumetric ranking consistent with expert assessment (Fig. 2).

Inter-rater agreement between the two independent experts was excellent (ICC > 0.94). The DynFGM framework achieved similarly high level of agreement with the experts for TAT (ICC = 0.92–0.97) and VAT (ICC = 0.82–0.90), approaching the level of inter-expert consensus. However, SAT agreement was notably more variable, falling within the moderate-to-good range (0.63–0.77). This suggests that while the framework accurately captures overall adipose trends, localized discrepancies remain at the subcutaneous boundaries where partial volume effects and expert extrapolation are most prevalent (Table 2).

Error Distribution and Bland–Altman Analysis

Volumetric error distributions for FCM and GMM were broad and skewed, whereas DynFGM errors were tightly distributed around zero (Fig. 3). Wilcoxon signed-rank tests confirmed that DynFGM errors were significantly lower than both baselines across all compartments ($p < 0.01$).

Bland–Altman analysis revealed minimal systematic bias, with limits of agreement (LoA) within < 10% of mean compartment volumes (Fig. 4). For TAT and SAT, the mean difference was near zero, with most cases falling within the 95% LoA. For the VAT compartment, a mean bias of -331.78 cm^3 was observed, but the LoA remained within clinically acceptable bounds for large-scale analysis. These findings support the framework's reliability for use in clinical research settings where manual segmentation is infeasible.

Physiological Plausibility in the Validation Cohort

In the large validation cohort ($n = 756$), DynFGM produced physiologically plausible adipose distributions (Table 3). Mean TAT volume was $11{,}457 \pm 3192$ cm^3, with SAT comprising $71.5 \pm 9.3\%$ and VAT $28.5 \pm 9.3\%$ of the total, respectively. Distributions remained symmetric (skewness: 0.23–0.34) with near-normal kurtosis (-0.41 to -0.04), suggesting consistent population-level characteristics across a heterogeneous population.

Parameter Stability and Sensitivity Analysis

The stability of the kurtosis-guided cluster selection was assessed by analyzing the distribution of excess kurtosis values across 343 slices. This analysis revealed that the majority of slices (67.1%, $n = 230$) possessed complex intensity distributions, requiring the higher granularity of a four-cluster model ($K = 4$) to resolve anatomical interfaces, while the $K = 2$ and $K = 3$ models were assigned to 12.5% and 20.4% of slices, respectively. Crucially, the empirical thresholds (-0.5 and 0.5) used for this adaptive selection were found to be strategically positioned in low-density regions of the kurtosis distribution. Only a small fraction of slices fell within a ± 0.15 perturbation margin of the high (3.5%) or low (9.6%) thresholds, indicating that for much of the volume (86.9%), the cluster assignment was decisive and inherently stable (Fig. 5).

To quantify the impact of these empirical choices on the final volumetric output, we conducted a perturbation analysis on a representative subset of participants ($n = 5$). The SAT/VAT separation logic demonstrated exceptional stability, with a mean VAT volume change of only 1.48% (range 0.00–4.91%) even when its core divisor was varied three-fold. The framework was also robust to changes in the model selection logic; perturbing the high and low kurtosis thresholds across a wide range resulted in only moderate volumetric shifts of 7.65% and 8.30%, respectively. This analysis confirms that the framework's quantitative outputs are not sensitive to the precise values of its parameters, validating the methodological robustness of the DynFGM approach (Fig. 6).

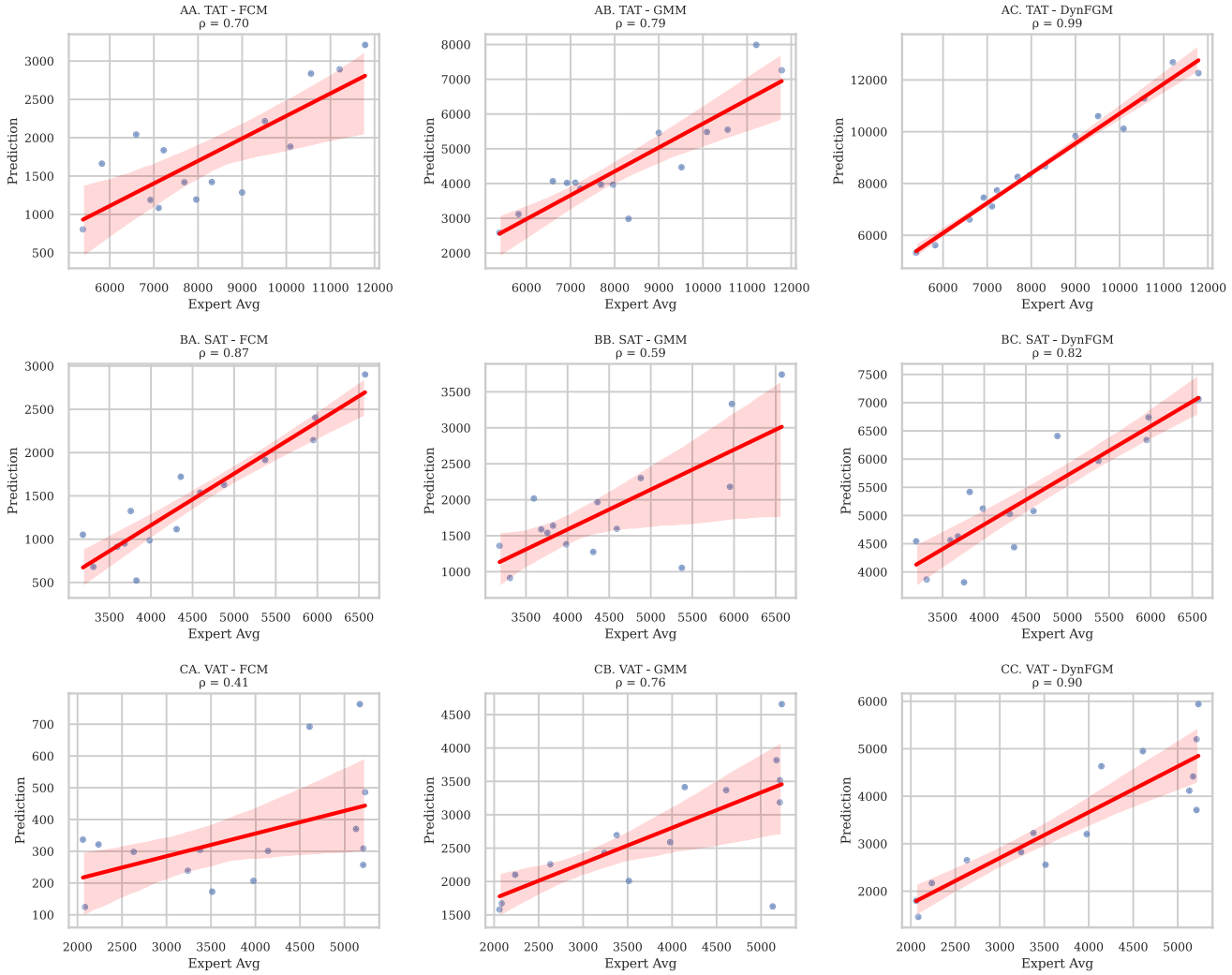

Fig. 2 Comparative correlation analysis of automated segmentation methods. Scatter plots illustrating the agreement between expert-consensus annotations (Expert Avg) and predictions generated by three unsupervised methods: fuzzy *C*-means (FCM), Gaussian mixture models (GMM), and the proposed Dynamic Fuzzy-Gaussian Modeling (DynFGM). Each plot includes a linear regression line (red) and a 95% confidence interval. Spearman's rank correlation coefficients (ρ) are provided for total adipose tissue (TAT), subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT). DynFGM demonstrates the highest orrelation across all compartments, specifically achieving $\rho = 0.99$ for TAT and significantly outperforming baseline methods in the challenging VAT compartment ($\rho = 0.90$)

Computational Performance and Failure Analysis

DynFGM completed fully automated segmentation in 13.6 ± 1.3 s per participant on a standard CPU (Intel® Core™ i9, 3.0 GHz). This remains orders of magnitude faster than manual delineation (≈ 30 min). Segmentation anomalies were observed in 23 of 756 scans (3.0%) during visual review (consistent with the $n = 756$ cohort). These cases typically involved body mask leakage or ambiguous VAT boundaries in low-contrast slices (Fig. S1). Importantly, all anomalies fell within the 95% Bland–Altman LoA, indicating minimal impact on aggregate quantitative metrics.

Discussion

The findings of this study establish DynFGM as a statistically principled advancement in unsupervised abdominal adipose tissue segmentation. In the expert-annotated benchmark cohort, the framework reduced volumetric error by over 90% compared to fixed-parameter baselines, achieving spatial agreement (DSC = 0.94) and inter-rater reliability (ICC up to 0.97). By integrating slice-adaptive, kurtosis-driven cluster selection with a hybrid FCM-GMM architecture, DynFGM addresses the structural rigidity inherent to classical clustering methods while avoiding the dependence on large, annotated training sets and limited interpretability of contemporary deep learning (DL) approaches [2, 31].

Table 2 Reliability of adipose tissue segmentation: inter-expert agreement and DynFGM performance ($ICC_{2,1}$)

Comparison	TAT	SAT	VAT
Expert 1 vs Expert 2	0.95	0.94	0.96
DynFGM vs Expert 1	0.92[a]	0.63[b]	0.90[a]
DynFGM vs Expert 2	0.97[a]	0.77[b]	0.82[b]

This table presents the intraclass correlation coefficients ($ICC_{2,1}$) evaluating the reliability and reproducibility of the DynFGM framework compared to manual rater consensus. Reliability interpretation based on standard clinical benchmarks: < 0.50 (poor), 0.50–0.75 (moderate), 0.75–0.90 (good), and > 0.90 (excellent)

ICC intraclass correlation coefficient (Model 2,1: two-way random effects, absolute agreement, and single rater/measurement); *TAT* total adipose tissue; *SAT* subcutaneous adipose tissue; *VAT* visceral adipose tissue

[a]Strong to excellent reliability in automated quantification for total and visceral compartments

[b]Moderate to good reliability observed in SAT separation, reflecting inherent variability in subcutaneous boundary definitions

The central methodological contribution of DynFGM lies in its dynamic adaptivity. Conventional unsupervised methods, including standard FCM and GMM, rely on a static number of clusters (K), rendering them sensitive to the pronounced intensity heterogeneity characteristic of multi-slice abdominal MRI [2, 11]. Our sensitivity analysis confirmed that the kurtosis-based logic for dynamically selecting the cluster number is robust. The empirical thresholds are strategically positioned in low-density regions of the data's distribution, ensuring a decisive cluster assignment for over 86.9% of slices. This intrinsic calibration of segmentation granularity prevents over-segmentation in homogeneous regions while introducing additional degrees of freedom ($K = 4$) to resolve anatomically complex interfaces. As such, DynFGM represents a mathematically transparent alternative to the heuristic cluster-selection rules employed in earlier algorithmic frameworks [32, 33].

Positioning DynFGM within the broader medical image analysis landscape requires both supervised deep learning pipelines and conventional unsupervised segmentation approaches. Supervised convolutional neural network (CNN) architectures, including FatSegNet, fully convolutional networks, and nnU-Net variants, have demonstrated state-of-the-art spatial overlap metrics for abdominal adipose tissue segmentation [2, 15, 34–36]. However, these methods generally depend on large expertly annotated datasets, substantial computational infrastructure, and protocol-specific training conditions. Previous studies have also reported reduced generalizability when applied to anatomically or technically heterogeneous datasets [13, 37]. In contrast, DynFGM operates in a fully unsupervised and label-free manner, requiring no annotated training data while maintaining high spatial agreement and volumetric reliability. Although supervised CNNs may achieve higher Dice coefficients under optimized training conditions, DynFGM offers several practical advantages, including computational efficiency, interpretability, reduced infrastructure dependence, and robustness in settings where annotated datasets are unavailable.

Traditional unsupervised and algorithmic segmentation methods provide greater transparency than DL approaches but often lack sufficient adaptability to local anatomical variation. Spatial FCM approaches improve robustness to noise through neighborhood constraints but retain a globally fixed cluster number and remain sensitive to intensity inhomogeneity and initialization parameters [37]. Similarly, Expectation–Maximization Gaussian mixture models (EM-GMM)

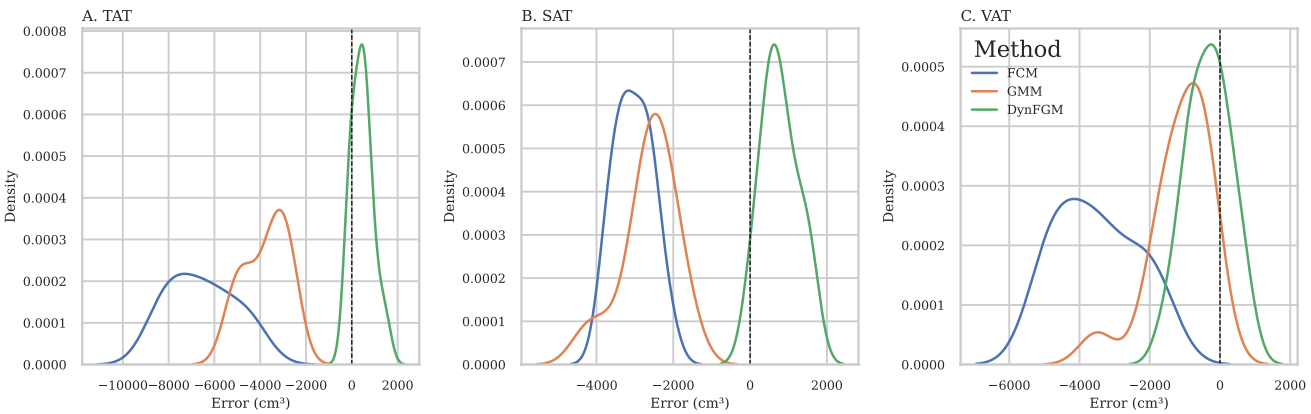

Fig. 3 Error distribution density for unsupervised segmentation methods. Kernel density estimation plots illustrate the distribution of volumetric errors (cm^3) for each method across the adipose depots. Panel A shows that for TAT, the proposed DynFGM method produces a narrow peak centered at zero, indicating minimal systematic bias and high precision. Panel B and C depict the error distributions for SAT and VAT, respectively. The DynFGM consistently exhibits the lowest error variance within the unsupervised domain. These distributions demonstrate that the proposed kurtosis-adaptive approach effectively mitigates the underestimation and overestimation errors observed in the baseline FCM and GMM models

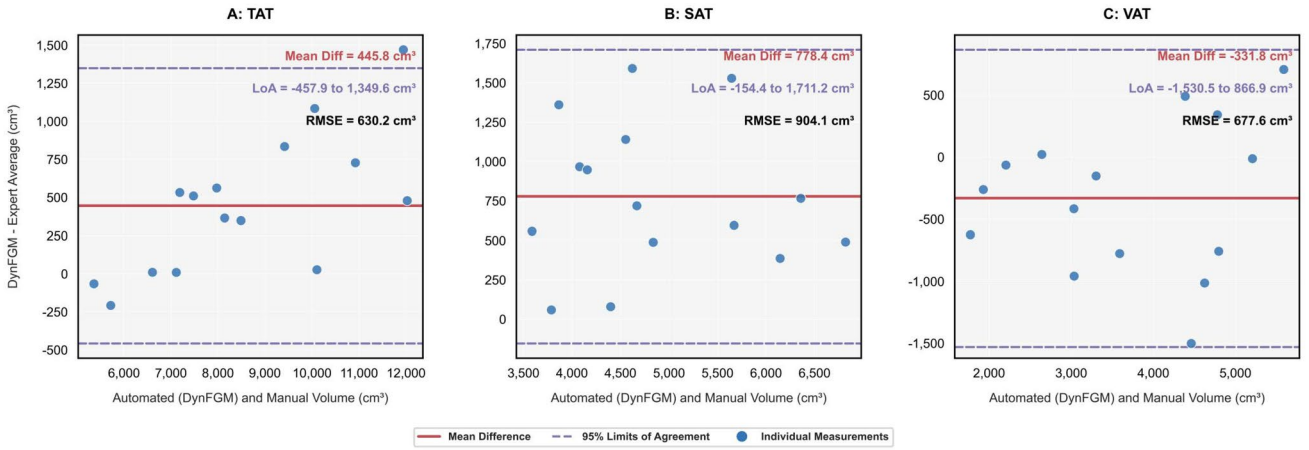

Fig. 4 Bland–Altman analysis comparing DynFGM with expert consensus. Bland–Altman plots were used to assess absolute agreement and potential systematic bias between the automated DynFGM framework and expert-consensus ground truth. **A, B** Agreement for total adipose tissue (TAT) and subcutaneous adipose tissue (SAT), respectively, showing that most data points fall within the 95% limits of agreement (±1.96 SD), with no evidence of proportional bias across the volumetric range. **C** Agreement for visceral adipose tissue (VAT), with a mean difference of − 331.78 cm^3. Although minor outliers are observed at the extreme ends of the volume distribution, the results confirm the overall reliability of the framework for population-level adipose tissue quantification

Table 3 Volumetric distribution and statistical characteristics of abdominal adipose tissue in the validation cohort ($n = 756$)

Adipose tissue	Mean ± SD (cm^3)	Median [IQR] (cm^3)	Range (cm^3)	Skewness	Kurtosis
TAT	11,457.3 ± 3192.3	11,390.2 [9057.8–13,495.9]	573.4–23,827.4	0.30	− 0.04
SAT	8168.5 ± 2433.5	8237.4 [6092.2–9884.9]	429.6–16,530.6	0.23	− 0.41
VAT	3288.8 ± 1422.4	3214.6 [2251.3–4233.9]	143.8–7766.9	0.34	− 0.27
SAT%[a]	71.5 ± 9.3	70.9 [63.9–78.9]	52.5–95.3		
VAT%[a]	28.5 ± 9.3	29.1 [21.1–36.1]	4.7–47.5		

The inclusion of skewness and kurtosis values characterizes the morphological distribution of these depots, while percentage values (SAT% and VAT%) illustrate the relative body composition profile of the cohort. Normal distribution benchmarks: skewness near 0 and kurtosis near 0 (excess kurtosis) indicate a relatively symmetric distribution of adipose volumes across the 756 participants

SD standard deviation, *IQR* interquartile range (25th–75th percentile), *TAT* total adipose tissue, *SAT* subcutaneous adipose tissue, *VAT* visceral adipose tissue

[a]Percentage calculated relative to the total adipose tissue (TAT) volume

provide an interpretable probabilistic framework for tissue classification but are susceptible to local minima and bias field distortions [38]. DynFGM addresses these limitations through slice-wise kurtosis-adaptive clustering, allowing the number of clusters ($K = 2$–4) to vary according to local image characteristics. The integration of ROF-TV denoising, N4 bias field correction, spatially constrained FCM initialization, and EM-GMM refinement enables robust tissue separation while preserving mathematical interpretability and eliminating dependence on labeled training data. To systematically outline the methodological and operational trade-offs, Table 4 evaluates the proposed work against leading automated pipelines across eight technical dimensions.

Moderate agreement was observed in the SAT compartment (ICC: 0.63–0.77), a finding attributable primarily to partial volume effects at high-contrast tissue boundaries, such as the dermal and fascial interfaces [18, 19]. Voxels in these regions contain mixed adipose and non-adipose signals. Whereas human raters frequently smooth these boundaries through spatial interpolation and anatomical judgment, DynFGM assigns boundary voxels strictly based on their statistical intensity profile. This fundamental distinction between subjective manual delineation and objective, intensity-driven clustering explains the comparatively lower inter-rater reliability observed for SAT. Nevertheless, Bland–Altman analysis demonstrated that the observed variability remained within clinically acceptable agreement limits, supporting the framework's reliability for large-scale adiposity quantification.

The practical utility of DynFGM is further underscored by its performance in the large validation cohort ($n = 756$), where the framework generated physiologically plausible

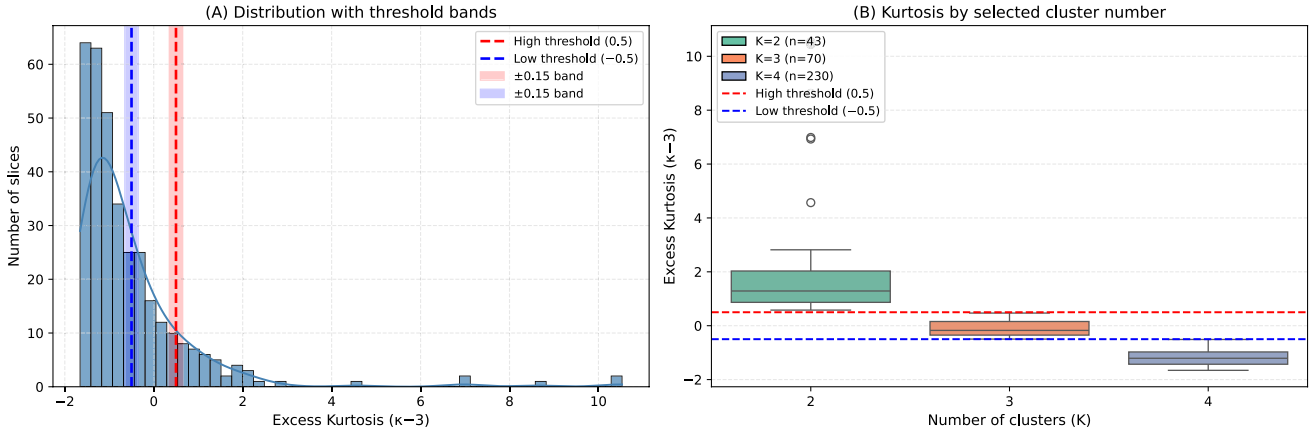

Fig. 5 Kurtosis distribution and adaptive cluster selection. **A** Global distribution of excess kurtosis ($k-3$) across 343 slices. Dashed vertical lines denote the empirical decision thresholds used for model selection ($K = 2,3,4$). The shaded bands represent a ± 0.15 perturbation margin employed in the sensitivity analysis; only a small minority of slices (3.5% and 9.6%) reside within these transition zones, indicating high algorithmic stability. **B** Relationship between excess kurtosis and the selected number of clusters. Boxplots demonstrate that the adaptive framework effectively partitions slices into statistically distinct groups. The majority of slices ($n = 230$) require the higher granularity of a four-cluster model ($K = 4$) to resolve anatomical complexity, while more homogeneous slices are correctly simplified to a two-cluster model ($K = 2$), thereby preventing over-segmentation

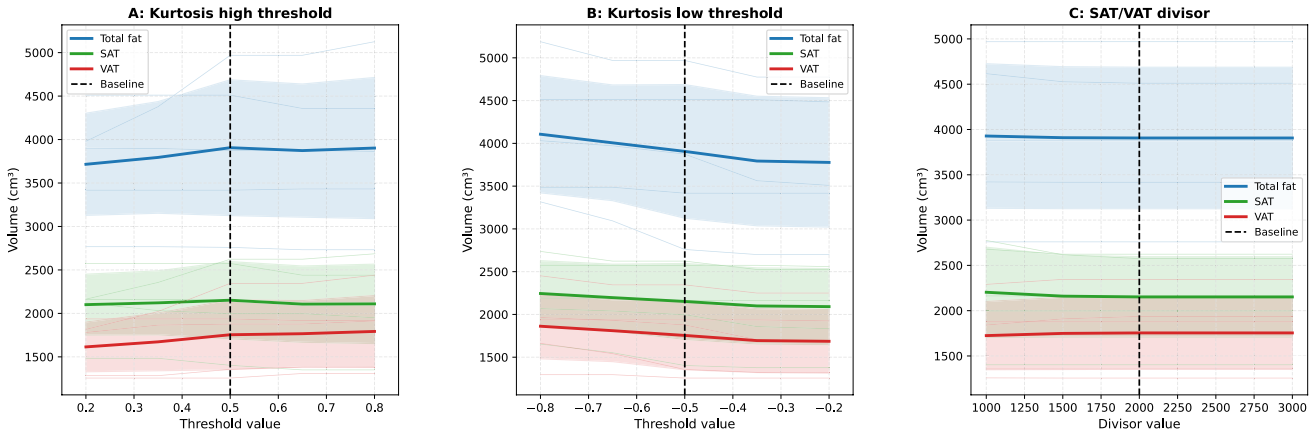

Fig. 6 Sensitivity analysis of empirical parameters. This figure illustrates the stability of volumetric outputs of adipose tissue under a wide range of parameter variations across a representative subset ($n = 5$). **A, B** Shifting the high and low kurtosis thresholds, which govern cluster number selection, results in minimal volumetric fluctuation, confirming that the model-selection logic is robust to threshold placement. **C** The SAT/VAT separation divisor demonstrates near-perfect stability across a threefold range (1000–3000), justifying the use of the empirical rule for anatomical delineation. Solid lines represent mean volumes, and shaded regions indicate the standard deviation across participants. The dashed vertical lines represent the baseline values used in the final DynFGM implementation

adiposity distributions with high computational efficiency (≈ 14 s per scan on a standard CPU). This throughput substantially exceeds the feasibility of manual delineation and supports potential integration into large-scale metabolic imaging studies and routine research workflows without requiring specialized AI hardware. Furthermore, the low segmentation anomaly rate (3.0%), combined with stable sensitivity analysis results, indicates that the framework is robust to moderate variations in anatomical complexity and image heterogeneity.

Limitations and Future Directions

Several limitations delineate the scope of this study. Quantitative benchmarking against expert annotations was performed on a relatively small manually labeled cohort ($n = 20$). Although robustness and scalability were demonstrated in a larger validation cohort ($n = 756$), we deliberately moderate claims of interchangeability with manual segmentation, and further validation across larger,

Table 4 Methodological and performance benchmarking matrix of automated abdominal adipose tissue segmentation pipelines

Models used	Imaging	Preprocessing engine	Approach type	Targets depots	Key performance metrics	Labels required?/Interpretability	Main strengths/limitations (vs DynFGM)
FatSegNet (CDFNet) [35]	Abdominal 3D Dixon MRI	Strict localization to Dixon fat/water contrasts	Supervised 2D CDFNets (localization + 3-view segmentation + aggregation)	SAT, VAT	DSC: SAT ≈ 0.975, VAT ≈ 0.850; ICC: 0.998–0.999	Yes (pixel-level annotations required); Low transparency (hidden dense network blocks)	Achieves higher SAT dice coefficients and high reliability, but requires a large, labeled training dataset and GPU resources; it is not unsupervised or adaptive to local slice anatomy
2D/3D FCN Architecture [14]	Water-fat Dixon MRI	Voxel resampling and bias field cleaning	Supervised DL (2D and 3D fully convolutional networks)	SAT, VAT,	DSC: SAT up to 0.992, VAT up to 0.988	Yes (massive multi-site expert tracing required); low transparency (latent layer representations)	Demonstrates high spatial accuracy on uniform datasets, but relies on large, labeled training sets and may exhibit performance variation under unseen scanner models
3D spatiotemporal CNN [2]	Multi-contrast MRI inputs	Cross-sequence spatiotemporal co-registration	Supervised DL (3D CNN utilizing multi-contrast longitudinal inputs)	SAT, VAT	Median DSC: SAT ≥ 0.992, VAT ≥ 0.976; ICC > 0.997	Yes (requires baseline cohorts; black-box model weights)	Provides high tracking speed and multi-contrast consistency across serial scans, but requires large baseline labeled cohorts, and evaluation was not conducted on matched data arrays
2D-CDFNet pipeline [39]	MRI sequence with T2 thresholding	Threshold-driven intensity filtering	Supervised 2D CDFNet	SAT, VAT	DSC: SAT ≈ 0.94, VAT ≈ 0.85	Yes (trained on manually labeled subset); low transparency (implicit neural weights)	Validated robust correlation against dual-energy X-ray absorptiometry but assumes a highly rigid protocol and fixed positioning
Spatial fuzzy *C*-means (sFCM) [37]	Multi-slice abdominal T1 MRI	Standard intensity mapping	Spatial unsupervised clustering	SAT, VAT	DSC: SAT ≈ 0.89, VAT ≈ 0.81	No labels; high transparency (fuzzy membership matrices)	Reduces local noise variations via spatial neighborhood constraints but relies on a globally static *K* and is sensitive to intense localized B1 field bias fields

Table 4 (continued)

Models used	Imaging	Preprocessing engine	Approach type	Targets depots	Key performance metrics	Labels required?/Interpretability	Main strengths/limitations (vs DynFGM)
Expectation-Maximization GMM [38]	Multi-echo Dixon sequences	Multi-echo water-fat separation	Parametric mixture modeling	SAT, VAT	Volumetric correlation ≈ 0.86 vs. manual tracing	No labels; high transparency (Gaussian probability densities)	Provides a probabilistic framework for partial volume estimations but can be sensitive to local minima initialization traps and local field distortions
Level set/local binary fitting [40]	Abdominal 1.5T MRI scans	Manual/heuristic contour seed initialization	Deformable contours & energy minimization	SAT, VAT	DSC for boundary contouring ≈ 0.88	No labels; high transparency (implicit contour energy functionals)	Delineates continuous physiological boundaries such as the abdominal wall but involves higher computational overhead per scan and requires manual heuristic tuning for contour forces
Markov random field clustering [41]	High-resolution T1 MRI	Baseline scale normalization	Contextual spatial field modeling	SAT, VAT	DSC: SAT ≈ 0.91, VAT ≈ 0.84	No labels; high transparency (explicit spatial interaction parameters)	Preserves contextual topological continuity across adjacent axial slices but increases computational complexity and relies on a static cluster count when slice-wise metrics shift
DynFGM *(proposed work)*	Abdominal Dixon MRI	ROF-TV denoising + N4 bias field correction + CLAHE	Unsupervised algorithmic (Kurtosis-driven adaptive cluster mapping, K=2–4 + FCM-initialized GMM-EM)	SAT, VAT, TAT	DSC ≈0.94 ± 0.03; ICC up to 0.97; >90% volumetric MAE reduction vs fixed-K baselines	No labels; high interpretability (explicit per-slice statistics, cluster counts, Gaussian components)	Bridges traditional rule-based and supervised DL: label-free and interpretable with adaptive complexity yet approaching supervised dice

DSC dice similarity coefficient, *ICC* intraclass correlation coefficient, *DL* deep learning, *CNN* convolutional neural network, *FCN* fully convolutional network, *CDFNet* competitive dense fully convolutional network, *RF* radiofrequency, *SAT* subcutaneous adipose tissue, *VAT* visceral adipose tissue, *TAT* total adipose tissue, *MAT* marrow adipose tissue, *IMAT* intermuscular adipose tissue, *ROF-TV* Rudin–Osher–Fatemi total variation, *CLAHE* contrast limited adaptive histogram equalization

independently annotated datasets is required to establish clinical-grade reliability. Additionally, all participants were overweight Asian adults; evaluation across populations with differing body compositions and ethnic backgrounds is necessary to confirm global generalizability. Future investigations will also explore integration of spatial regularization or hybrid statistical-learning strategies to further minimize boundary leakage.

Conclusions

DynFGM bridges the methodological gap between classical unsupervised clustering and modern deep learning by combining statistical adaptivity with probabilistic rigor. By dynamically calibrating segmentation granularity via slice-wise kurtosis, the framework achieved strong agreement with expert annotations, including high spatial overlap (DSC 0.94; ICC 0.82–0.97) and robust volumetric reliability (ICC up to 0.97), while substantially reducing volumetric error by over 90% compared with conventional FCM and GMM approaches. The framework further demonstrated stable performance and physiologically plausible adipose tissue distributions across a large validation cohort ($n = 756$), with low technical failure rates and efficient CPU-based processing, supporting its suitability for large-scale population imaging studies without reliance on specialized AI infrastructure.

Compared with conventional unsupervised clustering methods that rely on fixed cluster assumptions, DynFGM dynamically adapts segmentation complexity according to slice-specific anatomical heterogeneity while preserving mathematical interpretability and eliminating dependence on labeled training data. Although supervised deep learning frameworks may achieve higher Dice similarity metrics under optimized training conditions, DynFGM provides a transparent, computationally efficient, and label-free alternative for settings where annotated datasets are limited or unavailable.

Overall, the combination of interpretability, adaptive statistical modeling, computational efficiency, and label-free operation positions DynFGM as a scalable framework for automated adipose tissue quantification and metabolic phenotyping. Future work should focus on external validation across multi-center and multi-ethnic cohorts, evaluation under heterogeneous MRI acquisition protocols, and incorporation of spatial regularization or anatomical priors to further improve boundary delineation in anatomically complex regions. In addition, the proposed kurtosis-adaptive framework may be extended to other biomedical image segmentation applications characterized by tissue heterogeneity and limited labeled training data.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10278-026-02046-3.

Author Contribution Asefa Adimasu Taddese: conceptualization, methodology, software, and writing—original draft. Joshua D.K. Bernal: methodology, validation, formal analysis, and writing—original draft. Chit K. Leung: data curation, investigation, and writing—review and editing. Dominique Wang: visualization and software. Angus P. Yu: resources and supervision. Zuneera Umair: formal analysis and validation. Bjorn T. Tam: conceptualization, supervision, project administration, funding acquisition, and writing—review & editing. All authors have read and approved the final manuscript.

Funding Open access funding provided by Hong Kong Baptist University Library.

Data Availability The source code for the Dynamic Fuzzy-Gaussian Modeling (DynFGM) framework, including the kurtosis-adaptive clustering logic and pre-processing pipeline, will be made publicly available on GitHub upon publication. To ensure long-term reproducibility, a stable version of the code and a sample anonymized dataset used for validation will be archived on Zenodo with a registered Digital Object Identifier (DOI). Underlying clinical trial data are available from the corresponding author upon reasonable request and subject to institutional data privacy restrictions.

Declarations

Ethics Approval The MRI datasets analyzed in this study were secondary data obtained from a previously conducted clinical trial (ClinicalTrials.gov ID: NCT04887454). All data were fully anonymized prior to analysis. In accordance with institutional and local regulatory guidelines, the reuse of anonymized secondary data for methodological development and algorithm validation did not require additional ethical approval. All original data collection procedures were performed in accordance with the Declaration of Helsinki.

Consent to Participate For the original clinical trial from which this secondary data was derived, informed consent was obtained from all individual participants included in the study.

Consent for Publication Consent to Publish is not applicable, as the manuscript contains no individual person's data in any form (including individual details, images, or videos).

Competing interests The authors declare no competing interests.

References

1. Storz C, Heber SD, Rospleszcz S, Machann J, Sellner S, Nikolaou K, et al. The role of visceral and subcutaneous adipose tissue measurements and their ratio by magnetic resonance imaging in subjects with prediabetes, diabetes and healthy controls from a general population without cardiovascular disease. The British journal of radiology. 2018;91(1089):20170808.
2. Kafali SG, Shih S-F, Li X, Kim GHJ, Kelly T, Chowdhury S, et al. Automated abdominal adipose tissue segmentation and volume quantification on longitudinal MRI using 3D convolutional neural networks with multi-contrast inputs. Magnetic Resonance Materials in Physics, Biology and Medicine. 2024;37(3):491-506.
3. Alberti KGMM, Zimmet P, Shaw J. Metabolic syndrome—a new world-wide definition. A consensus statement from the international diabetes federation. Diabetic medicine. 2006;23(5):469–80.
4. Barrera F, Labbé P, Liberona J, Arrese M, Irarrázabal P, Tejos C, et al. Quantification of visceral adipose tissue using magnetic resonance imaging compared with anthropometry, in type 2 diabetic patients. Revista Medica de Chile. 2012;140(12):1535-43.
5. Cesaro A, De Michele G, Fimiani F, Acerbo V, Scherillo G, Signore G, et al. Visceral adipose tissue and residual cardiovascular risk: a pathological link and new therapeutic options. Frontiers in cardiovascular medicine. 2023;10:1187735.
6. Kullberg J, Ahlström H, Johansson L, Frimmel H. Automated and reproducible segmentation of visceral and subcutaneous adipose tissue from abdominal MRI. International journal of obesity. 2007;31(12):1806-17.
7. Ogier AC, Hostin M-A, Bellemare M-E, Bendahan D. Overview of MR image segmentation strategies in neuromuscular disorders. Frontiers in Neurology. 2021;12:625308.
8. Kullberg J, Von Below C, Lönn L, Lind L, Ahlström H, Johansson L. Practical approach for estimation of subcutaneous and visceral adipose tissue. Clinical physiology and functional imaging. 2007;27(3):148-53.
9. Hu HH, Chen J, Shen W. Segmentation and quantification of adipose tissue by magnetic resonance imaging. Magnetic Resonance Materials in Physics, Biology and Medicine. 2016;29(2):259-76.
10. Bouazizi K, Zarai M, Dietenbeck T, Aron-Wisnewsky J, Clément K, Redheuil A, et al. Abdominal adipose tissue components quantification in MRI as a relevant biomarker of metabolic profile. Magnetic Resonance Imaging. 2021;80:14-20.
11. Zhou A, Murillo H, Peng Q. Novel segmentation method for abdominal fat quantification by MRI. Journal of Magnetic Resonance Imaging. 2011;34(4):852-60.
12. Peng Q, McColl RW, Ding Y, Wang J, Chia JM, Weatherall PT. Automated method for accurate abdominal fat quantification on water-saturated magnetic resonance images. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2007;26(3):738-46.
13. Devi B, Misbha D. Automated Visceral Adipose Tissue Segmentation and Quantification From Abdominal Mri Using an Enhanced U-Net And Region-Growing. South Eastern European Journal Of Public Health. 2024:2827–44.
14. Langner T, Hedström A, Mörwald K, Weghuber D, Forslund A, Bergsten P, et al. Fully convolutional networks for automated segmentation of abdominal adipose tissue depots in multicenter water–fat MRI. Magnetic resonance in medicine. 2019;81(4):2736-45.
15. Kafali SG, Shih S-F, Li X, Chowdhury S, Loong S, Barnes S, et al., editors. 3D neural networks for visceral and subcutaneous adipose tissue segmentation using volumetric multi-contrast MRI. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021: IEEE.
16. Renard F, Guedria S, Palma ND, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. Scientific Reports. 2020;10(1):13724.
17. Positano V, Cusi K, Santarelli MF, Sironi A, Petz R, DeFronzo R, et al. Automatic correction of intensity inhomogeneities improves unsupervised assessment of abdominal fat by MRI. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2008;28(2):403-10.
18. Tohka J. Partial volume effect modeling for segmentation and tissue classification of brain magnetic resonance images: A review. World journal of radiology. 2014;6(11):855.
19. Kway YM, Thirumurugan K, Tint MT, Michael N, Shek LP-C, Yap FKP, et al. Automated segmentation of visceral, deep subcutaneous, and superficial subcutaneous adipose tissue volumes in MRI of neonates and young children. Radiology: Artificial Intelligence. 2021;3(5):e200304.
20. Pang Y, Li L, Hu W, Peng Y, Liu L, Shao Y. Computerized Segmentation and Characterization of Breast Lesions in Dynamic Contrast-Enhanced MR Images Using Fuzzy c-Means Clustering and Snake Algorithm. Computational and Mathematical Methods in Medicine. 2012;2012(1):634907.
21. Adhikari SK, Sing JK, Basu DK, Nasipuri M. Conditional spatial fuzzy C-means clustering algorithm for segmentation of MRI images. Applied soft computing. 2015;34:758-69.
22. Shiwaku K, Anuurad E, Enkhmaa B, Kitajima K, Yamane Y. Appropriate BMI for Asian populations. The Lancet. 2004;363(9414):1077.
23. Siu PM, Leung CK, Bernal JD, Yu AP, Recchia F, Tam BT, et al. Once and thrice weekly interval training in adults with central obesity: a randomized controlled trial. Nature Communications. 2026.
24. Kuk JL, Church TS, Blair SN, Ross R. Does measurement site for visceral and abdominal subcutaneous adipose tissue alter associations with the metabolic syndrome? Diabetes care. 2006;29(3):679-84.
25. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic resonance imaging. 2012;30(9):1323-41.
26. Heo Y-C, Kim K, Lee Y. Image denoising using non-local means (NLM) approach in magnetic resonance (MR) imaging: a systematic review. Applied Sciences. 2020;10(20):7028.
27. Kanakaraj P, Yao T, Cai LY, Lee HH, Newlin NR, Kim ME, et al. DeepN4: learning N4ITK bias field correction for T1-weighted images. Neuroinformatics. 2024;22(2):193-205.
28. Fortin M, Omidyeganeh M, Battié MC, Ahmad O, Rivaz H. Evaluation of an automated thresholding algorithm for the quantification of paraspinal muscle composition from MRI images. Biomedical engineering online. 2017;16(1):61.
29. Haugen C, Lysne V, Haldorsen I, Tjora E, Gudbrandsen OA, Sagen JV, et al. MRI adipose tissue segmentation and quantification in R (RAdipoSeg). Diabetology & Metabolic Syndrome. 2022;14(1):146.
30. Song Y, Ji Z, Sun Q, editors. An extension Gaussian mixture model for brain MRI segmentation. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.
31. Schneider D, Eggebrecht T, Linder A, Linder N, Schaudinn A, Blüher M, et al. Abdominal fat quantification using convolutional networks. European Radiology. 2023;33(12):8957-64.
32. Chen Y, Cheng N, Cai M, Cao C, Yang J, Zhang Z. A spatially constrained asymmetric Gaussian mixture model for image segmentation. Information Sciences. 2021;575:41-65.

33. Schwab M, Mayr A, Haltmeier M, editors. Deep Gaussian mixture model for unsupervised image segmentation. International conference on machine learning, optimization, and data science; 2024: Springer.
34. Chaitanya K, Erdil E, Karani N, Konukoglu E. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. Medical image analysis. 2023;87:102792.
35. Estrada S, Lu R, Conjeti S, Orozco-Ruiz X, Panos-Willuhn J, Breteler MM, et al. FatSegNet: A fully automated deep learning pipeline for adipose tissue segmentation on abdominal dixon MRI. Magnetic resonance in medicine. 2020;83(4):1471-83.
36. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods. 2021;18(2):203-11.
37. Chuang K-S, Tzeng H-L, Chen S, Wu J, Chen T-J. Fuzzy c-means clustering with spatial information for image segmentation. Computerized Medical Imaging and Graphics. 2006;30(1):9-15.
38. Greenspan H, Ruf A, Goldberger J. Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. IEEE Transactions on Medical Imaging. 2006;25(9):1233-45.
39. Wu T, Estrada S, Gils Rv, Su R, Jaddoe VWV, Oei EHG, et al. Automated Deep Learning–Based Segmentation of Abdominal Adipose Tissue on Dixon MRI in Adolescents: A Prospective Population-Based Study. American Journal of Roentgenology. 2024;222(1):e2329570.
40. Li C, Kao CY, Gore JC, Ding Z, editors. Implicit Active Contours Driven by Local Binary Fitting Energy. 2007 IEEE Conference on Computer Vision and Pattern Recognition; 2007 17–22 June 2007.
41. Held K, Kops ER, Krause BJ, Wells WM, Kikinis R, Muller-Gartner HW. Markov random field segmentation of brain MR images. IEEE Transactions on Medical Imaging. 1997;16(6):878-86.