European Journal of Radiology Open 16 (2026) 100749

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Development and validation of a fully automated transformer-based 3D framework for pancreatic fat quantification in pancreatic steatosis

You Pang [a], Gary Ge Ren [a], Raymond Shing Yan Tang [b], Winnie Chu [b], Chileka Chiyanika [a,b,*]

[a] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
[b] Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Prince of Wales Hospital, Sha Tin, Hong Kong, China

HIGHLIGHTS

- Developed an automated pancreatic segmentation and PDFF quantification pipeline.
- TransUNet achieved the highest segmentation accuracy using combined water-only and fat-only MRI.
- AI-based method strongly correlates with manual volumetric PDFF quantification, outperforming ROI-based methods.

ARTICLE INFO

Keywords:
Pancreas
Adipose Tissue
Deep Learning
Magnetic Resonance Imaging
Adiposity

ABSTRACT

Purpose: To develop and validate a fully automated TransUNet-based framework for 3D pancreatic segmentation and volumetric fat quantification on PDFF MRI. Specifically, to evaluate the model's performance against state-of-the-art CNN architectures and assess its clinical reproducibility compared to manual reference methods.
Methods: This retrospective study involved 140 adults with metabolic dysfunction-associated steatotic liver disease who underwent 3.0 T multi-echo mDIXON MRI. A TransUNet-based model was developed using a 5-fold cross-validation approach, integrating convolutional and transformer layers to capture global and local features. Various architectures (UNet, nnUNet) and multiple input combinations of water, fat, and R2* maps were systematically compared. Segmentation performance was primarily assessed using the Dice Similarity Coefficient (DSC), with the Jaccard index, precision, recall, and 95th-percentile Hausdorff distance additionally used for network comparison. Agreement of pancreatic fat quantification across measurement methods was assessed using Bland-Altman analysis and intraclass correlation coefficients (ICC).
Results: TransUNet achieved the highest segmentation accuracy (DSC: 85.46%, IQR: 80.76–86.24%), outperforming nnUNet (84.24%) and UNet (71.89%). Optimal performance was reached using water-only and fat-only input series. The AI-based volumetric method demonstrated strong agreement with manual whole-organ PDFF ($r = 0.866$, $p < 0.001$) with a minimal mean bias (1.20). Conversely, 2D ROI methods significantly underestimated pancreatic fat ($6.08 \pm 2.74\%$) compared to both manual ($21.95 \pm 6.14\%$) and AI-based ($20.72 \pm 6.13\%$) volumetric assessments ($p < 0.001$).
Conclusion: TransUNet provides accurate, reproducible 3D pancreatic segmentation and fat quantification. By capturing the entire organ volume, this automated framework overcomes the sampling bias inherent in traditional 2D ROI methods, offering a fast and reliable biomarker for pancreatic steatosis.

1. Introduction

Fatty pancreas is a significant predictor of beta cell dysfunction, and type 2 diabetes mellitus [1,2]. Although clinically significant, accurate non-invasive quantification remains challenging. Routine practice depends on small region-of-interest (ROI) sampling, which is prone to sampling bias due to the marked spatial heterogeneity of pancreatic fat [3,4]. Furthermore, 2D ROI methods significantly underestimate total

* Correspondence to: Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Room Y938, 11 Yuk Choi Road, Hung Hom, Hong Kong, China.
E-mail addresses: you.pang@polyu.edu.hk (Y. Pang), gary-ge.ren@polyu.edu.hk (G.G. Ren), raymondtang@cuhk.edu.hk (R.S.Y. Tang), winniechu@cuhk.edu.hk (W. Chu), chileka.chiyanika@polyu.edu.hk (C. Chiyanika).

https://doi.org/10.1016/j.ejro.2026.100749
Received 25 November 2025; Received in revised form 24 March 2026; Accepted 11 April 2026
Available online 17 April 2026

fat compared to whole-organ volumetric assessments [5]. While manual whole-organ segmentation is more accurate, it is labour-intensive and poorly reproducible, limiting its use in clinical workflows [6].

The technical challenge in automated pancreas delineation stems from the organ's small size, variable morphology, and low soft-tissue contrast against adjacent viscera [3]. Early automated attempts using atlas-based or statistical shape models achieved only moderate accuracy and required expert interaction [7]. Though Deep Convolutional Neural Networks (CNNs) like UNet and nnUNet have set strong baselines on contrast-enhanced CT, their reliance on local convolutional kernels often fails to capture the global contextual information necessary for the ill defines boundaries of the pancreas on MRI [7,8]

Recent hybrid architectures, such as TransUNet, address these limitations by integrating Transformers with CNNs. By leveraging self-attention mechanisms, these models aggregate global spatial dependencies, which are critical for distinguishing intrapancreatic fat from surrounding visceral adipose tissue [9]. Despite these advances, fully automated 3D frameworks for volumetric fat quantification on proton density fat fraction (PDFF) MRI, the gold standard for fat measurement remain scarce and typically require manual correction [10].

Therefore, this study aims to bridge this gap by developing and validating a fully automated TransUNet-based framework for 3D pancreatic segmentation and volumetric fat quantification. Specifically, the objectives are: (1) To develop a transformer-based 3D algorithm for automated pancreas segmentation and PDFF-based fat measurement; (2) To evaluate its performance against state-of-the-art CNN architectures using standardised segmentation metrics; (3) To assess the clinical reproducibility of the automated volumetric measurements in comparison to manual reference methods.

2. Materials and methods

2.1. Subjects

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the local institution (CREC reference number 2023.263). All participants provided informed written consent.

This retrospective study involved Chinese adults (aged 18–65) who underwent chemical shift encoded MRI (mDIXON) to diagnose or confirm metabolic dysfunction-associated steatotic liver disease (MASLD) following a positive fatty liver screening using the controlled attenuation parameter. Inclusion criteria were intrahepatic triglyceride content measured by MRI-proton density fat fraction (PDFF) $\geq$ 5% and at least one metabolic risk factor as defined by Rinella et al. [11]. Exclusion criteria included alcohol consumption $>$ 30 g/day for men or 20 g/day for women, MRI contraindications, and a history of liver transplantation or resection. To ensure findings were representative of metabolic-associated steatotic liver disease MASLD and to maintain consistency with the primary study, subjects with hepatocellular or pancreatic carcinoma, other malignancies (unless in remission for over five years), or pancreatic lesions were excluded. These exclusions prevent the structural and metabolic alterations associated with malignancy from confounding the analysis and ensure the accuracy of the automated segmentation baseline. Obesity was defined as BMI $\geq$ 25 kg/m^2, and morbid obesity as BMI $\geq$ 40 kg/m^2, or BMI $\geq$ 32 kg/m^2 plus T2DM or two obesity-related co-morbidities for the Asian population [12]. T2DM was defined by fasting glucose levels (venous plasma $\geq$ 7.0 mmol/L; whole blood; $\geq$ 6.1 mmol/L), 2-hour post-glucose load levels (venous plasma $\geq$ 11.1 mmol/L; whole blood $\geq$ 10.0 mmol/L), or an HbA1c level $\geq$ 6.5% (48 mmol/mol), with all study participants having a confirmed diagnosis for at least six months prior to enrolment [12].

2.2. Chemical shift encoded MRI data acquisition

All subjects underwent MRI using a Philips Achieva 3.0 T scanner (Philips Medical System, Best, The Netherlands) with a 16-channel SENSE-XL-Torso array coil after an 8-hour fast. High-resolution water-fat separation was performed using a 3D spoiled multi-echo mDIXON sequence. Detailed imaging parameters are summarised in Table 1.

Image reconstruction was performed online using the vendor-supplied Philips mDIXON product implementation (Philips Healthcare, Best, The Netherlands). This utilises a multi-echo chemical-shift water-fat separation algorithm incorporating a multi-peak spectral model of fat and T2* correction to enhance accuracy and sensitivity [13].

Two 15-second breath-hold acquisitions were used to cover the entire abdomen. This broad coverage was chosen because the primary clinical objective of the original scan was to evaluate both the liver (for MASLD) and the pancreas, ensuring that any anatomical variations in organ position or ectopic fat distribution were captured within the same co-registered series. Furthermore, this wide field-of-view enabled the comprehensive measurement of visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) across the whole abdomen, providing a complete profile of abdominal adiposity. Minor variations in in-plane resolution and slice number were present across the dataset due to patient-specific adjustments and scanner calibration, resulting in anisotropic voxel spacings addressed during preprocessing.

To ensure the integrity of these volumetric data, all eligible examinations underwent a secondary image quality and usability screening before downstream analysis (Fig. 1). This step was applied after clinical eligibility was confirmed and was performed by two abdominal imaging specialists with over 15 and 25 years of experience, respectively. MRI series were excluded if they exhibited significant motion or breathing artefacts that were sufficient for reliable delineation, or if file-related issues were identified (e.g., format or encoding errors that prevented successful data loading or standardised preprocessing). It is acknowledged that the exclusion of low-quality images may not fully reflect real-world clinical practice; this is addressed as a study limitation.

2.3. Measurement of pancreatic proton density fat fraction

2.3.1. Region of Interest (ROI) method

Measurements were performed by two abdominal imaging specialists with over 15 and 25 years of experience, respectively, and were blinded to clinical data. Pancreatic PDFF was measured using the fat fraction image series. Three regions of interest (ROIs), each 1 cm^2, were drawn on the head, body, and tail of the pancreas, avoiding the pancreatic duct and splenic vein, using the Philips IntelliSpace Portal version R3.0-SP15 (Philips Healthcare, Best, The Netherlands). This was performed three times on any slice where the pancreas was clearly visible. The mean PDFF from these three ROIs was averaged to determine the mean pancreatic fat fraction.

2.3.2. Manual Whole-organ Volumetric Pancreatic Fat Quantification (Slice-by-Slice Method)

For manual whole-organ volumetric quantification, the pancreas was identified in each abdominal MR image using anatomical landmarks (e.

Table 1
Summary of MRI Acquisition Parameters.

Parameter	Value
Sequence Type	3D spoiled multi-echo mDIXON
Repetition Time (TR)	5.7–5.9 ms
Echo Time (TE) / Echo Spacing	1.2–1.4 ms / 1.0–1.2 ms (6 echoes)
Flip Angle	3°
FOV	375 × 313 × 219 mm
Acquisition Matrix	128 × 117
Reconstructed Slice Thickness	3.0 mm
Acceleration (SENSE)	2

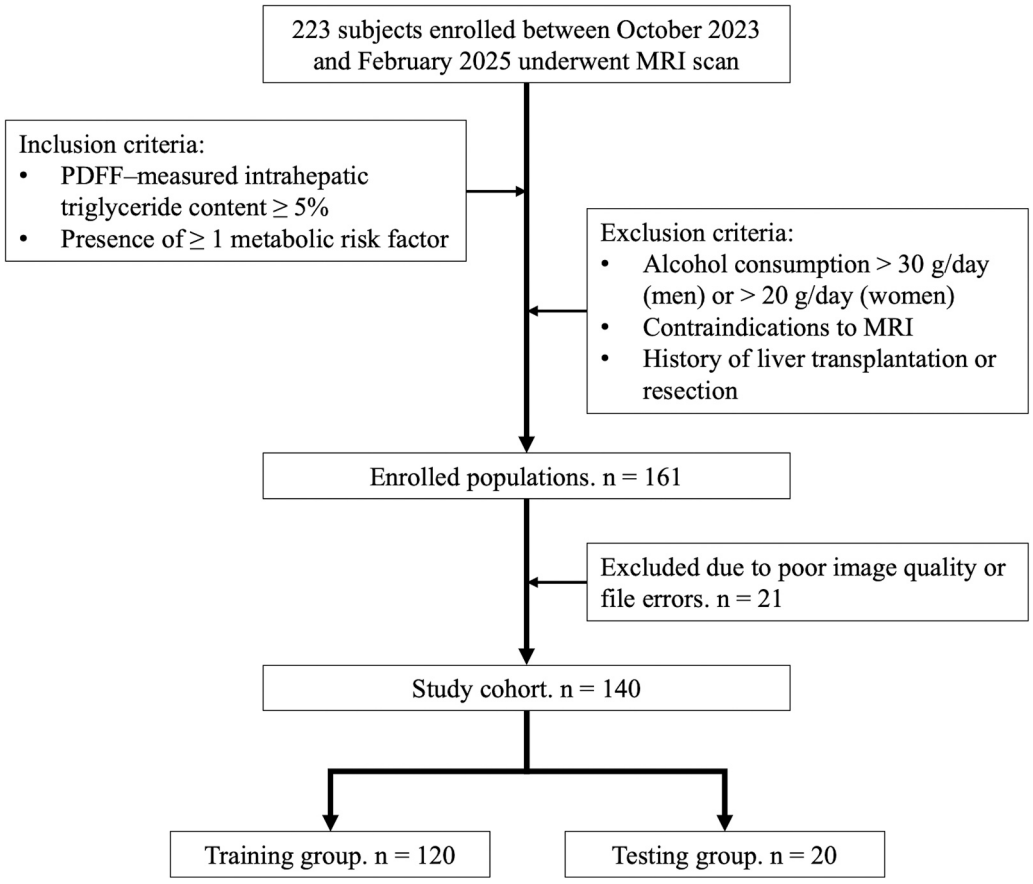

Fig. 1. Flowchart of the study subject recruitment process. Eligible patients were first identified using predefined clinical inclusion/exclusion criteria. Their MRI scans were then screened for image quality to exclude inadequate images before forming the final analysis cohorts.

g., splenic vein, superior mesenteric artery, inferior vena cava, aorta, duodenum). Using the smoothed polygon ROI tool in the Philips IntelliSpace Portal version R3.0-SP15 (Philips Healthcare, Best, The Netherlands), the pancreas was carefully delineated on every slice where

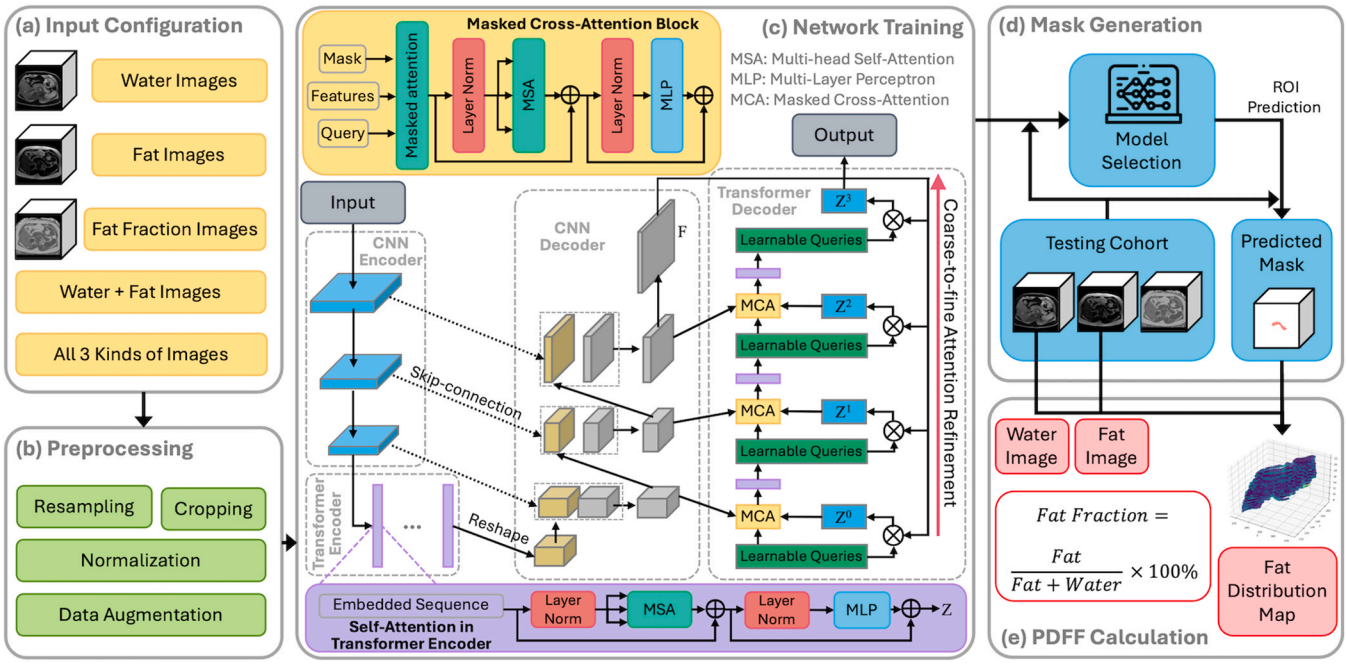

Fig. 2. Overview of the TransUNet-based pipeline for pancreatic segmentation and PDFF quantification. (a-b) Data preparation: 5 input configurations are tested. (c) The TransUNet model was trained and validated to optimize segmentation performance. (d) The best-performing model was applied to the testing cohort to generate predicted pancreas masks. (e) Pancreatic PDFF was then calculated voxel-wise using the water images, fat images, and predicted masks.

pancreatic tissue was visible. This process resulted in a set of contiguous polygonal regions that together encompassed the entire pancreas across all slices.

Within these manually outlined regions of the fat fraction images, the software automatically extracted the fat fraction (PDFF) values from the fat fraction images provided by the MRI vendor. The final whole-organ PDFF was calculated as the mean of all pancreatic voxels across slices, ensuring that each voxel contributed equally.

This approach ensures that slices with smaller pancreatic surface area do not disproportionately influence the overall PDFF measurement, as could occur if per-slice mean PDFF values were averaged together. By calculating the mean PDFF across all pancreatic voxels, we avoid bias and achieve a true volumetric assessment of pancreatic fat content.

2.4. AI-based 3D pancreatic segmentation and automated fat fraction quantification

We developed an AI-based framework to automatically segment the pancreas on multiparametric MRI and subsequently quantify whole-pancreas fat content. To examine the contribution of different MRI contrasts, five input configurations were tested, including both single-modality and multimodality combinations of water, fat, and fat fraction images (Fig. 2(a)).

Before model training, all images underwent a standardized preprocessing pipeline (Fig. 2(b)). Briefly, the image volumes were resampled to a common voxel spacing, cropped to a fixed 3D region covering the pancreas, and normalized on a per-case basis to reduce inter-scan variability. Standard on-the-fly data augmentation was applied during training to improve model robustness.

The segmentation model was based on TransUNet, which combines convolutional layers and transformer modules (Fig. 2(c)). In this framework, convolutional layers capture local anatomical details, whereas transformer modules provide broader contextual information across the slice. This combined design was used to improve pancreatic delineation, particularly given the irregular shape of the pancreas and its often indistinct boundaries on MRI. The network was trained from scratch without pre-trained weights.

Manual whole-pancreas annotations served as the reference standard and were generated slice by slice in ITK-SNAP by an abdominal MRI specialist and a radiologist with over 20 years of combined experience. The dataset was divided into training and testing groups at a ratio of 6:1, with cases in the testing group drawn from a different time period to provide temporal separation from the training data. Within the training group, 80% of cases were used for model training and 20% for validation. Five TransUNet models were trained using identical data splits and training settings, each corresponding to one input configuration, to enable fair comparison. In addition, UNet and nnUNet were evaluated as comparator architectures.

During inference, the trained model generated a 3D pancreatic mask for each examination using a sliding-window strategy (Fig. 2(d)). The final model used for testing was selected according to validation performance. The predicted pancreatic mask was then applied to the paired water-only and fat-only images to derive a voxel-wise fat fraction map within the segmented pancreas (Fig. 2(e)). Specifically, for each voxel within the pancreas region, the relative contribution of the fat content was estimated based on the corresponding fat-only and water-only image intensities, thereby generating a voxel-level fat fraction map. This procedure provides a spatially resolved quantitative characterization of fat distribution across the pancreas. The mean value across all voxels within the segmented pancreatic volume was used as the automated whole-pancreas PDFF for each subject.

3. Statistical analysis

Statistical analyses were conducted using Python (v3.7) and SPSS version 28. Segmentation model performance was evaluated by comparing model-predicted pancreatic masks with manual annotations using the Dice Similarity Coefficient (DSC).

For benchmarking, UNet and nnUNet models were trained alongside TransUNet and evaluated using DSC. nnUNet employed its default self-configuration, while UNet used the same hyperparameters as TransUNet; preprocessing and augmentation were identical across models. Performance was first assessed by 5-fold cross-validation, reporting mean ± standard deviation (SD) DSC for each model and input configuration. The best-performing model from each group was then evaluated case-wise on the test set, with results summarized by median, interquartile range (IQR), and mean ± SD. Statistical comparisons used the Friedman test for global differences, followed by Nemenyi post-hoc tests with a critical difference (CD) approach. Wilcoxon signed-rank tests with Holm correction were additionally applied to control for multiple comparisons. A two-sided $p < 0.05$ was considered significant.

Descriptive statistics summarised clinical, biochemical, and imaging characteristics. Continuous variables were reported as mean ± standard deviation (SD), while categorical variables were presented as percentages. Pancreatic PDFF values from different measurement methods were reported as median (IQR).

Differences in mean PDFF among ROI-based, manual whole-organ, and AI-automated methods were compared using ANOVA with Bonferroni post hoc analysis. Agreement and consistency between methods were evaluated using Pearson correlation coefficients (r) with 95% confidence intervals and p-values. Correlation strength was classified as very weak ($r < 0.3$), weak ($0.3 \leq r < 0.5$), moderate ($0.5 \leq r < 0.7$), or strong ($r \geq 0.7$) [14]. Bland–Altman analysis was performed to assess agreement between AI-based and manual whole-organ volumetric PDFF measurements. Statistical significance was set at $p < 0.05$.

Table 2
Characteristics of the subjects based on clinical, serological, and MRI analysis.

Variable	Mean ± SD
Age (years)	51.35 ± 11
Sex (Male/Female), (%)	56/43
BMI (kg/m^2)	35.89 ± 3.5
With obesity (%)	64.4
With morbid obesity, (%)	35.6
Hypertension (%)	55.7
Dyslipidaemia (%)	75.7
T2DM (%)	64.3
Total Cholesterol (mmol/l)	4.41 ± 1.1
Triglycerides (mmol/l)	1.62 ± 0.97
HDL (mmol/l)	1.23 ± 0.32
LDL (mmol/l)	2.48 ± 0.79
Fasting Plasma Glucose (mmol/l)	6.13 ± 1.50
HbA1c (%)	6.56 ± 1.18
ALP (IU/l)	75.98 ± 38.96
ALT (IU/l)	47.76 ± 32.79
AST (IU/l)	33.53 ± 16.55
GGT (IU/l)	50.19 ± 40.51
Liver PDFF (%)	12.80 ± 7.16
Pancreatic PDFF (%)- ROI	6.08 ± 2.74
Pancreatic PDFF (%)- Whole manual	21.95 ± 6.14
Pancreatic PDFF (%)- AI	20.72 ± 6.13

Note: obesity means BMI $\geq$ 25 kg/m^2, and morbid obesity means BMI $\geq$ 40 kg/m^2, or BMI $\geq$ 32 kg/m^2 plus Type 2 Diabetes or two obesity-related co-morbidities for the Asian population [12]. Abbreviations: SD, standard deviation; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate transaminase; BMI, body mass index; GGT, gamma-glutamyl transferase; HbA1c, glycosylated haemoglobin; HDL, high density lipoprotein; low density lipoprotein; PDFF, proton density fat fraction, T2DM, type 2 diabetes mellitus; ROI, region of interest, AI, artificial intelligence.

4. Results

4.1. Fat quantification

Table 2 presents the baseline characteristics of the study cohort with MASLD (n = 140). The mean pancreatic PDFF value was significantly lower when measured using the ROI method (6.08 ± 2.74%) compared to both the manual whole-organ volumetric method (21.95 ± 6.14%) and the AI-based segmentation method (20.72 ± 6.13%), both $p < 0.001$. There was no statistically significant difference between the manual whole-organ PDFF volumetric and AI-based PDFF quantification methods (21.95 ± 6.14% vs. 20.72 ± 6.13%, $p = 0.056$).

There was a strong correlation between the manual whole-organ PDFF volumetric and AI-based PDFF measurements ($r = 0.866$, 95% CI: 0.815–0.904, $p < 0.001$). In contrast, neither the AI-based PDFF nor the manual whole-organ PDFF volumetric measurements correlated with the ROI-based PDFF (AI-based: $r = 0.072$, 95% CI: −0.103–0.243, $p = 0.417$; manual: $r = 0.060$, 95% CI: −0.115–0.231, $p = 0.504$).

Bland-Altman analysis showed an absolute mean difference of 0.99 (95% CI: 0.49–1.49) between the AI-based and manual whole-organ volumetric PDFF measurements, with 95% limits of agreement ranging from −4.87–6.86. No evidence of proportional bias was observed ($p = 0.977$). In comparison, both the AI-based and manual whole-organ volumetric methods showed large absolute mean differences from the ROI method of −14.96 (95% CI: −16.02 to −13.90) and −15.95 (95% CI: −17.00 to −14.90), respectively, with significant proportional bias in both cases ($p < 0.001$), suggesting that AI-based whole-organ volumetric PDFF is a clinically viable and superior replacement for traditional ROI-based sampling. Further details are provided in Fig. 3.

4.2. Inter-rater and Intra-rater Reliability

The reliability of the manual reference methods was high. For the ROI method, the intraclass correlation coefficients (ICC) for absolute agreement ranged from 0.877 to 0.969. The inter-rater correlation coefficient between the two abdominal imaging specialists was 0.824, indicating strong consistency in manual pancreatic fat quantification.

4.3. Segmentation performance

Three deep-learning architectures were compared using combined water-only and fat-only images as input. In 5-fold cross-validation, TransUNet achieved the highest mean DSC (83.03%±0.65%), outperforming nnUNet (81.70%±0.64%) and UNet (70.12%±0.75%). We selected the best model from each of these three architectures for case-wise comparison. Case-wise DSC are summarized as follows: UNet (Median 71.89%, IQR 67.45–74.89; Mean 71.23 ± 5.25), nnUNet (Median 84.24%, IQR 78.38–85.62; Mean 82.62 ± 4.73), and TransUNet (Median 85.46%, IQR 80.76–86.24; Mean 83.83 ± 4.51). The Friedman test confirmed significant overall differences (χ^2=30.70, $p < 0.001$). On average, TransUNet outperformed nnUNet by 1.21% and UNet by 12.60% in DSC. To provide a more comprehensive evaluation beyond DSC, we additionally report other metrics in Table 3, which further characterize boundary accuracy and error patterns across models. Notably, even modest DSC improvements can translate into more stable organ delineation and downstream volumetric measurements, particularly in thin or anatomically ambiguous regions (e.g., the pancreatic tail), where small boundary deviations may disproportionately affect fat quantification. Details are shown in Figs. 4 and 5(d) – (f).

Given the superior performance of TransUNet, we further compared its performance across different input modality combinations. In 5-fold cross-validation, the Water + Fat configuration achieved the highest DSC (83.03%±0.65%), outperforming single-modality inputs (Fat: 80.69%±0.57%; Water: 80.59%±0.75%; Fat Fraction: 78.36%±0.70%) and the all-modalities configuration (79.31%±0.76%). Case-wise analysis confirmed the advantage of Water + Fat (Median 85.46%, IQR 80.76–86.24), with significant differences over all other settings according to the Friedman test (χ^2=20.52, $p < 0.001$) and post-hoc comparisons (Table 4, Fig. 5(a) – (c)). A 3D visualization of TransUNet-derived pancreatic PDFF is shown in Fig. 6. Note that DSC comparisons are reported for pancreas segmentation under different input combinations, whereas PDFF quantification is a downstream step computed within the segmentation mask on the PDFF map derived from the Dixon acquisition.

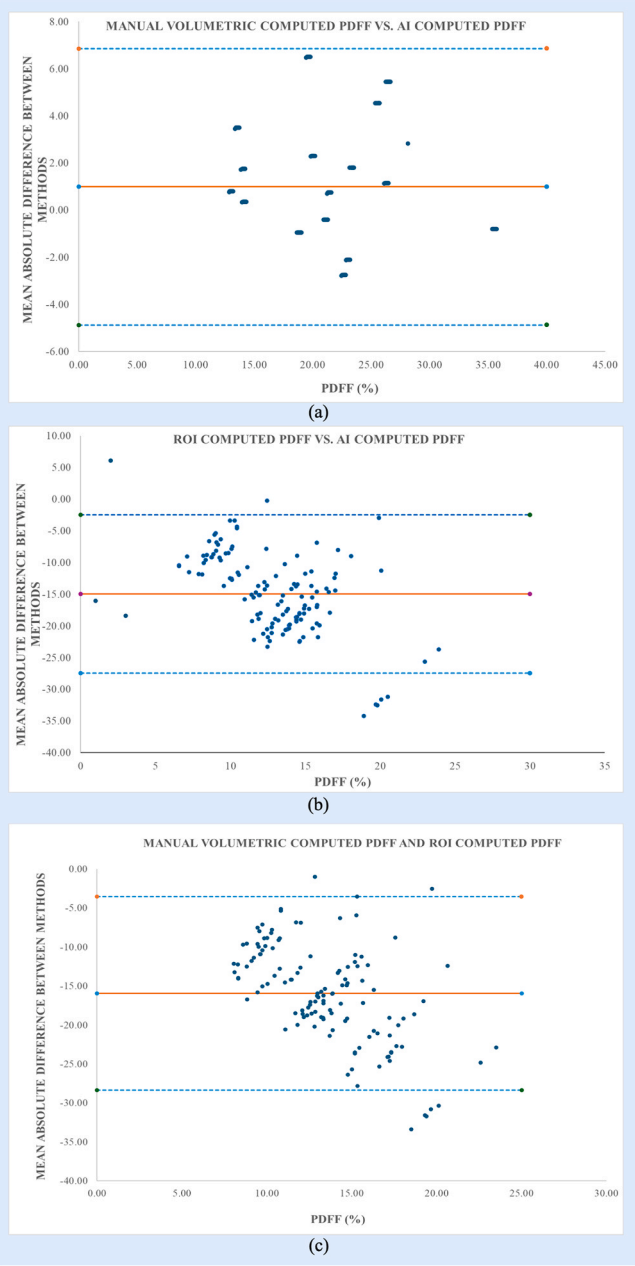

Fig. 3. Bland-Altman analysis of PDFF measurements. (a) Illustrates the consistency between the outcome from the manual-whole organ calculated volumetric Pancreatic PDFF method and the AI-computed PDFF method, with no proportional bias detected ($p = 0.977$). (b) demonstrates the consistency between the outcome from the ROI-based Pancreatic PDFF method and the AI-computed PDFF method, where a proportional bias was identified ($p < 0.001$). (c) Shows the consistency between the outcome from the manual-whole organ calculated volumetric Pancreatic PDFF method and the ROI-based Pancreatic PDFF method, with a proportional bias observed ($p < 0.001$).

Table 3
Quantitative segmentation performance evaluation and comparison with other models.

Methods	DSC (%)	Jaccard (%)	Precision (%)	Recall (%)	HD 95 (mm)
UNet	71.23 ± 5.25	55.53 ± 6.32	78.27 ± 12.41	67.61 ± 10.14	16.49 ± 6.69
nnUNet	82.62 ± 4.73	70.64 ± 6.79	85.78 ± 10.47	80.85 ± 7.78	7.66 ± 4.75
TransUNet	83.83 ± 4.51	72.47 ± 6.52	87.87 ± 7.11	81.02 ± 8.35	5.36 ± 1.54

Note: DSC, Dice Similarity Coefficient; Jaccard, Jaccard Index; HD 95, 95th percentile Hausdorff Distance; %, percentage; mm, millimetres.

4.4. Sub-analyses in terms of diabetes, sex, and obesity statuses

Subgroup analyses were performed to evaluate the robustness of the automated framework across different clinical phenotypes. When stratified by diabetes status, no significant differences in pancreatic PDFF were observed between those with and without T2DM across all three quantification methods ($p > 0.05$; Table 5). Similarly, sex-based analysis showed consistent fat quantification across all methods ($p > 0.05$; Table 6), suggesting the AI model's performance is not influenced by gender-specific anatomical variations.

However, significant discrepancies emerged when evaluated by obesity status (Table 7). For the ROI-based method, a significant difference in the pancreatic fat was observed between the obesity and morbid obesity groups (5.58% ± 2.73% vs. 6.98% ± 2.58%, $p = 0.017$). In contrast, both the manual whole-organ ($p = 0.244$) and AI-based ($p = 0.222$) methods showed no statistically significant differences between these groups. This suggests that as BMI increases, and pancreatic boundaries become less defined, small ROI sampling is highly susceptible to sampling bias and spatial heterogeneity. Conversely, the AI-driven whole-organ segmentation provided a more stable and representative assessment, effectively distinguishing intrapancreatic fat from the increased visceral adiposity typical of morbidly obese patients.

5. Discussion

This study validated a fully automated TransUNet-based framework for 3D pancreatic segmentation and fat quantification on PDFF MRI. Our findings demonstrate that this automated approach effectively addresses the limitations of manual assessment, offering a reproducible alternative for whole-organ volumetric analysis.

In agreement with prior studies [15,16], these results show that conventional 2D ROI methods significantly underestimate pancreatic fat compared to volumetric techniques. This discrepancy stems from the marked spatial heterogeneity of pancreatic fat as 2D ROIs fail to capture the total organ volume leading to substantial sampling bias. The subgroup analysis further highlighted this limitation. While ROI-based measurements showed statistically significant differences between obesity and morbid obesity groups ($p = 0.017$), both AI and manual whole-organ methods remained stable ($p > 0.05$). This suggests that as BMI increases and visceral fat infiltrates the peripancreatic space, small ROI sampling becomes increasingly unreliable. In contrast, the AI-driven volumetric approach remains robust regardless of obesity status, providing a more stable and representative assessment of total pancreatic steatosis.

A primary technical challenge in this domain is distinguishing intrapancreatic fat from adjacent visceral adipose tissue. While the close anatomical relationship between the pancreas and surrounding viscera risks inadvertent inclusion of non-parenchymal fat, our model trained on expert-delineated landmarks, demonstrated high specificity. The strong correlation between manual and AI-based measurements, with a mean bias of only 0.99%, supports the clinical reliability of the

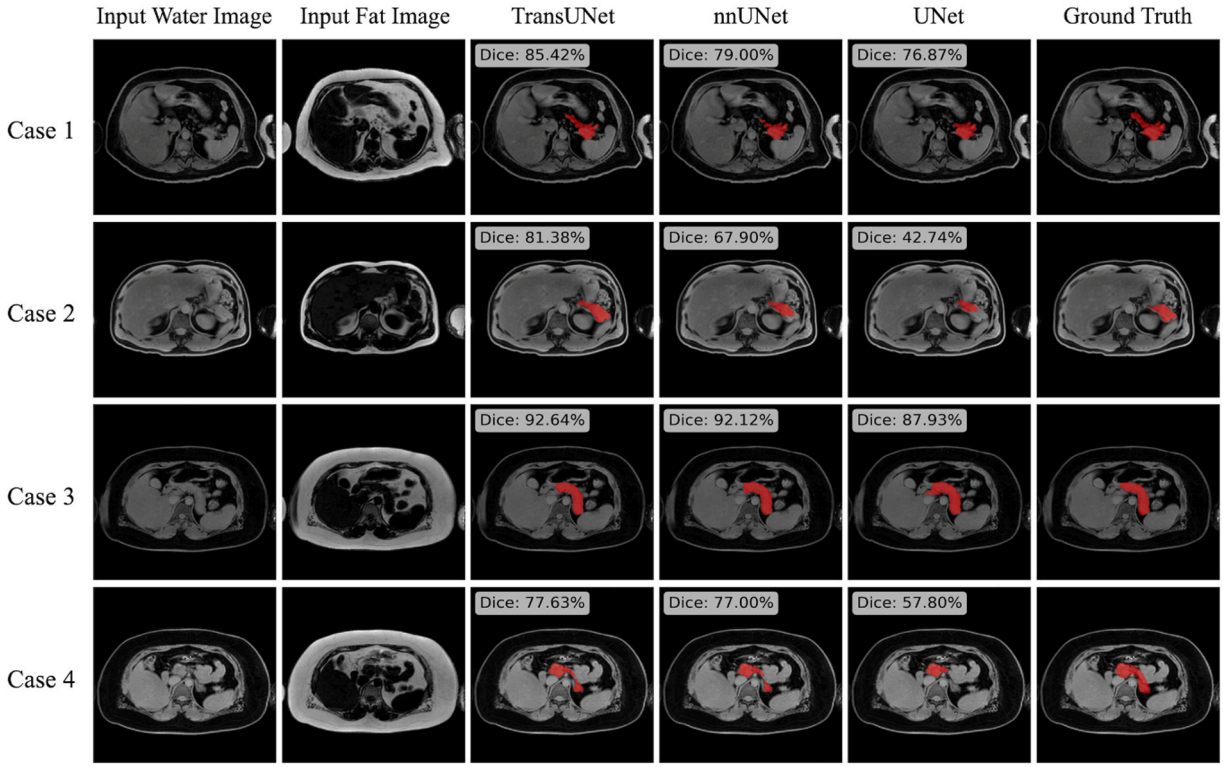

Fig. 4. Qualitative comparison of pancreas segmentation results from three deep learning models using combined Water and Fat image inputs. Five representative test cases are shown in rows (Case 1–4). Segmentation outputs are overlaid in red. TransUNet and nnUNet demonstrate better conformity with the anatomical boundaries of the pancreas, particularly in regions with indistinct edges (e.g., tail regions in Case 1), while UNet shows evident under-segmentation and discontinuities, consistent with its lower quantitative performance.

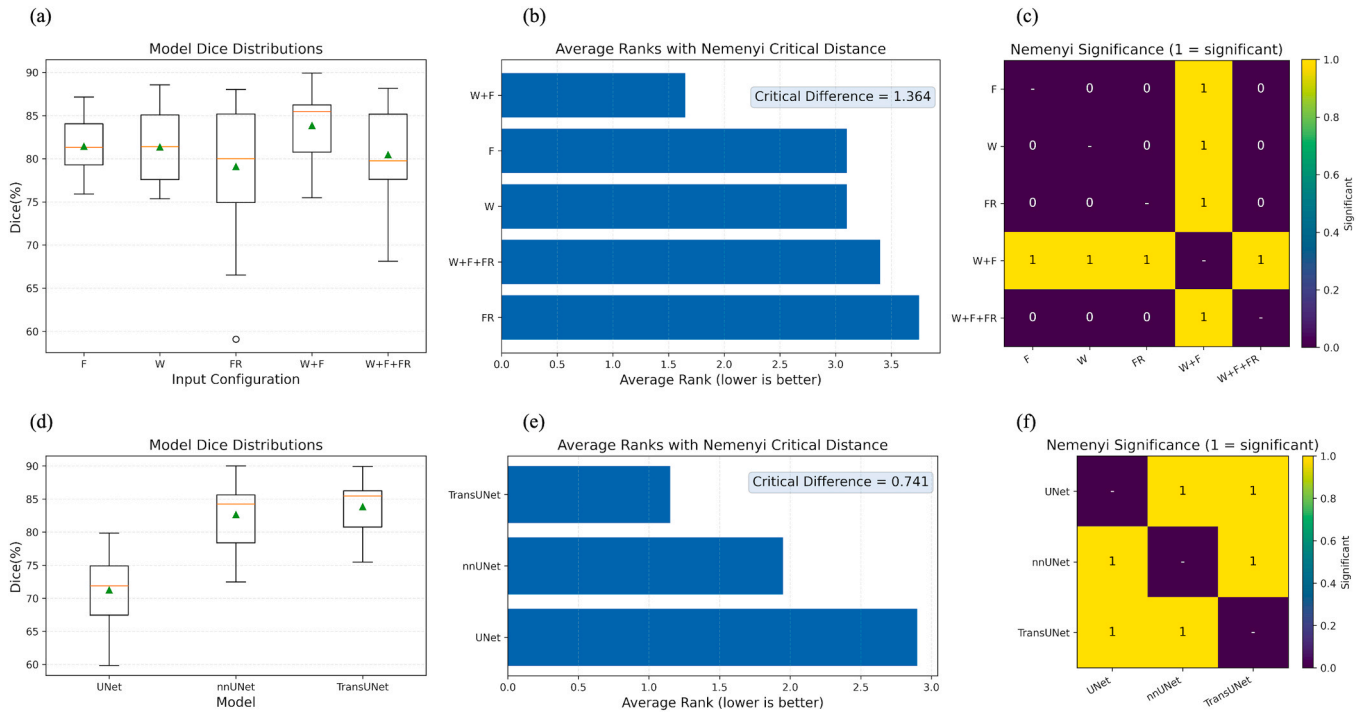

Fig. 5. Statistical comparison of segmentation performance using the Friedman test and post-hoc analysis. (a, d) DSC distributions across different input configurations and deep-learning models, with a line at the median and a triangle at the mean, and outliers are represented by hollow dots. (b, e) Average ranks of the five input configurations (with CD=1.364) and three deep-learning models (with CD=0.741) are indicated for the Nemenyi test. (c, f) Pairwise significance matrix of models based on the Nemenyi test (value=1 denotes significant difference at α=0.05). F, Fat-only images; W, Water-only images; FR, Fat fraction images; W+F, Water and Fat images; W+F+FR, Water, Fat, and Fat fraction images.

Table 4
Comparison of TransUNet model performance across various input configurations.

TransUNet Input Configuration	Q1-DSC (%)	Median-DSC (%)	Q3-DSC (%)	IQR-DSC (%)	Mean ± SD (%)
Water-only images	77.59	81.32	85.06	7.47	81.34 ± 4.24
Fat-only images	79.28	81.32	84.05	4.77	81.43 ± 3.32
Fat Fraction-only images	74.93	80.02	85.17	10.25	79.08 ± 7.43
Water + Fat images	80.76	85.46	86.24	5.48	83.83 ± 4.51
All images	77.61	79.75	85.16	7.55	80.44 ± 5.68

Note: Q1, quartile 1; Q3, quartile 3; IQR, interquartile range; DSC, Dice Similarity Coefficient; %, Percentage; SD, standard deviation

automated approach. Crucially, this bias is well within the 10% intervals typically used for clinical steatosis grading, suggesting the AI-based method is unlikely to lead to diagnostic misclassification.

The TransUNet architecture achieved a DSC of 83.03%, performing at or above current benchmarks for MRI-based pancreatic segmentation [17]. However, it is important to note that because the AI was trained on manual expert annotations, the high concordance between the two likely reflects the model's success in replicating expert performance rather than providing a completely independent ground truth. While these results align with findings from Janssens et al. [18] regarding the potential of automated volumetric analysis, further external validation against independent datasets is necessary to confirm universal clinical utility.

Among the evaluated models, TransUNet's superior performance is likely due to its hybrid design, which captures both local convolutional features and global contextual dependencies which are essential for delineating the variable morphology of the pancreas [19]. It was found that multimodal inputs (water-only and fat-only) provided the best results. Although R2* mapping was initially considered, it was found to be less effective for segmentation due to susceptibility artefacts and lower signal-to-noise ratios [20,21]. Consequently, the study focused on high-contrast modalities to ensure robust and reproducible automated workflows.

Despite these promising results, this study has several limitations. First, as a single-centre study involving only Chinese adults with MASLD, the findings may not be fully generalisable to other ethnicities or scanner vendors. The reliance on a single scanner model and protocol introduces potential 'domain shift' issues; future work should incorporate multi-vendor datasets to ensure robustness. Second, our relatively small sample size and retrospective design preclude longitudinal assessment of disease progression. Third, pancreatic fat quantification lacked histopathological confirmation, and the framework was not tested on patients with focal lesions or malignancy. Finally, while the exclusion of low-quality images (e.g., those with motion artefacts or encoding errors) ensured high internal validity for model training, we acknowledge that this curated dataset may not fully reflect the technical challenges of real-world clinical practice. Consequently, the framework's performance on suboptimal scans remains to be established in more heterogeneous clinical cohorts.

In conclusion, this study demonstrates that a TransUNet-based volumetric framework can accurately replicate expert manual segmentation for pancreatic fat quantification while significantly outperforming traditional ROI-based methods in terms of stability and efficiency. By capturing the entire organ volume, the AI-based approach overcomes the sampling bias inherent in 2D techniques, particularly in patients with high BMI, where fat distribution is most heterogeneous. While further multicentre validation and histopathological correlation are required to establish absolute accuracy, these study findings highlight the potential of AI-driven PDFF MRI analysis to provide a reliable,

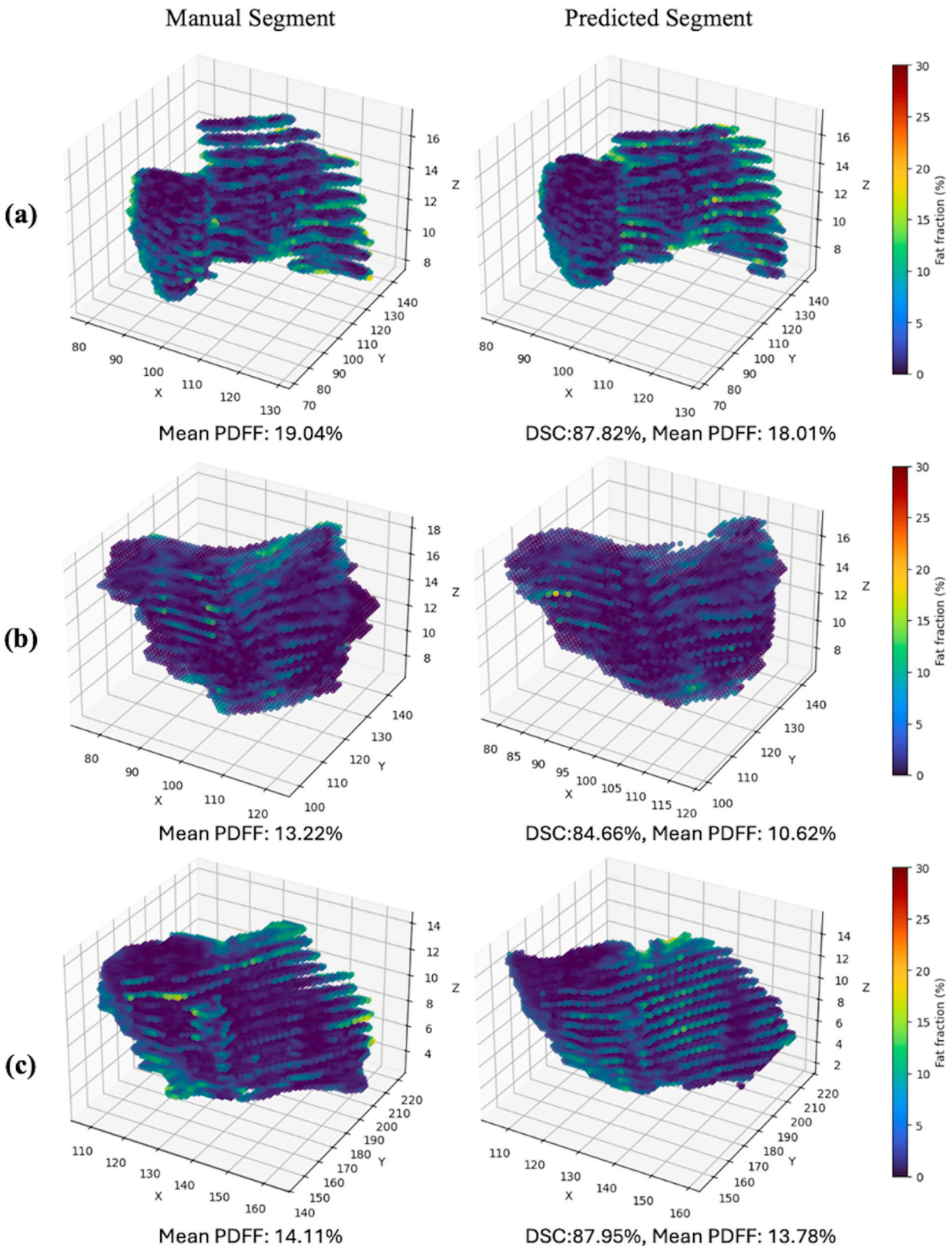

Fig. 6. 3D visualization of pancreatic fat distribution maps derived from manual and AI-predicted segmentations. Panels (a)-(c) show representative cases with voxel-wise fat fraction (%) overlaid on 3D reconstructed pancreatic masks. For each case, the left column presents the manual segmentation results, while the right column displays the predicted segmentations generated by the best TransUNet model. Mean PDFF values and DSC are reported for each case. Across all cases, the AI-predicted segmentation demonstrated close agreement with manual annotation, both in terms of spatial fat distribution and global PDFF quantification, supporting the reliability of the proposed automated approach.

Table 5
Characteristics of the subjects based on Diabetes status.

Variable	Without Diabetes (n = 50)	With Diabetes (n = 90)	p-value
Age (years)	49.06 ± 14.34	53.34 ± 8.23	0.116
BMI (kg/m^2)	34.83 ± 2.24	35.90 ± 3.95	0.283
Total Cholesterol (mmol/l)	4.77 ± 1.17	4.26 ± 1.01	0.059
Triglycerides (mmol/l)	1.46 ± 0.66	1.70 ± 1.12	0.811
HDL (mmol/l)	1.35 ± 0.36	1.19 ± 0.27	0.054
LDL (mmol/l)	2.80 ± 0.69	2.34 ± 0.82	**0.016**
Fasting Plasma Glucose (mmol/l)	5.28 ± 0.77	6.53 ± 1.60	**< 0.001**
HbA1c (%)	5.83 ± 0.56	6.94 ± 1.26	**< 0.001**
ALP (IU/l)	76.78 ± 21.97	76.72 ± 45.98	0.418
ALT (IU/l)	46.52 ± 27.43	47.80 ± 34.56	0.887
AST (IU/l)	32.13 ± 10.67	33.48 ± 16.73	0.606
GGT (IU/l)	60.57 ± 51.78	46.30 ± 34.16	0.225
Liver PDFF (%)	12.30 ± 6.84	13.24 ± 7.59	0.663
Pancreatic PDFF (%)- ROI	6.04 ± 3.40	6.10 ± 2.41	0.560
Pancreatic PDFF (%)- Whole manual	21.70 ± 5.53	22.66 ± 7.06	0.227
Pancreatic PDFF (%)- AI	20.29 ± 5.21	21.94 ± 7.33	0.326

Independent Samples Test was used. SD, standard deviation; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate transaminase; BMI, body mass index; GGT, gamma-glutamyl transferase; HbA1c, glycosylated haemoglobin; HDL, high density lipoprotein; low density lipoprotein; PDFF, proton density fat fraction, T2DM, type 2 diabetes mellitus; ROI, region of interest, AI, artificial intelligence. Significance is $p < 0.05$.

Table 6
Characteristics of the subjects based on sex.

Variable	Female (n = 50)	Male (n = 90)	p-value
Age (years)	51.05 ± 12.19	51.58 ± 10.18	0.841
BMI (kg/m^2)	36.71 ± 3.49	34.75 ± 3.38	0.184
Total Cholesterol (mmol/l)	4.71 ± 1.07	4.19 ± 1.04	**0.041**
Triglycerides (mmol/l)	1.63 ± 0.80	1.61 ± 1.08	0.930
HDL (mmol/l)	1.31 ± 0.35	1.17 ± 0.28	0.074
LDL (mmol/l)	2.58 ± 0.91	2.40 ± 0.70	0.345
Fasting Plasma Glucose (mmol/l)	6.38 ± 1.51	5.95 ± 1.49	0.225
HbA1c (%)	6.60 ± 1.09	6.53 ± 1.25	0.799
ALP (IU/l)	75.27 ± 28.91	76.49 ± 45.27	0.890
ALT (IU/l)	47.10 ± 30.55	48.24 ± 34.70	0.884
AST (IU/l)	36.17 ± 21.17	31.55 ± 11.93	0.289
GGT (IU/l)	44.79 ± 34.25	54.14 ± 44.54	0.340
Liver PDFF (%)	13.78 ± 6.17	12.08 ± 7.80	0.302
Pancreatic PDFF (%)- ROI	5.87 ± 2.90	6.24 ± 2.64	0.579
Pancreatic PDFF (%)- Whole manual	21.73 ± 6.78	22.12 ± 5.70	0.800
Pancreatic PDFF (%)- AI	21.31 ± 6.83	20.29 ± 5.61	0.503

Independent Samples Test was used. SD, standard deviation; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate transaminase; BMI, body mass index; GGT, gamma-glutamyl transferase; HbA1c, glycosylated haemoglobin; HDL, high density lipoprotein; low density lipoprotein; PDFF, proton density fat fraction, T2DM, type 2 diabetes mellitus; ROI, region of interest, AI, artificial intelligence. Significance is $p < 0.05$.

automated biomarker for the management of metabolic disease in clinical practice.

CRediT authorship contribution statement

Ren Ge Gary: Writing – review & editing, Software, Methodology, Conceptualization. **Tang Shing Yan Raymond:** Writing – review & editing, Supervision, Data curation, Conceptualization. **Pang You:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Chu Winnie:** Writing – review & editing, Supervision, Data curation, Conceptualization. **Chiyanika Chileka:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Table 7
Characteristics of the subjects based on Obesity status.

Variable	With Obesity (n = 90)	With Morbid obesity (n = 50)	p-value
Age (years)	54.75 ± 11.13	45.20 ± 7.71	**< 0.001**
BMI (kg/m^2)	28.61 ± 0.35	36.55 ± 2.82	**< 0.001**
Total Cholesterol (mmol/l)	4.57 ± 1.17	4.12 ± 0.85	**0.032**
Triglycerides (mmol/l)	1.66 ± 1.04	1.55 ± 0.83	0.310
HDL (mmol/l)	1.27 ± 0.36	1.17 ± 0.23	0.076
LDL (mmol/l)	2.58 ± 0.79	2.29 ± 0.78	0.066
Fasting Plasma Glucose (mmol/l)	6.02 ± 1.44	6.34 ± 1.62	0.204
HbA1c (%)	6.54 ± 1.14	6.60 ± 1.28	0.430
ALP (IU/l)	82.29 ± 44.81	65.05 ± 22.80	**0.018**
ALT (IU/l)	51.76 ± 36.14	40.85 ± 25.16	0.070
AST (IU/l)	35.05 ± 17.10	30.96 ± 15.57	0.156
GGT (IU/l)	56.64 ± 48.13	39.03 ± 17.72	**0.015**
Liver PDFF (%)	12.70 ± 7.71	12.99 ± 6.17	0.432
Pancreatic PDFF (%)- ROI	5.58 ± 2.73	6.98 ± 2.58	**0.017**
Pancreatic PDFF (%)- Whole manual	21.29 ± 5.81	22.32 ± 6.35	0.244
Pancreatic PDFF (%)- AI	20.01 ± 5.55	21.12 ± 6.45	0.222

Independent Samples Test was used. SD, standard deviation; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate transaminase; BMI, body mass index; GGT, gamma-glutamyl transferase; HbA1c, glycosylated haemoglobin; HDL, high density lipoprotein; low density lipoprotein; PDFF, proton density fat fraction, T2DM, type 2 diabetes mellitus; ROI, region of interest, AI, artificial intelligence. Significance is $p < 0.05$.

Ethical statement

We declare that this study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee on 6 December 2023 (REF NO. 2023.563). All participants provided informed written consent.

Funding

This work was supported by the Hong Kong Polytechnic University, grant number: 1-BD3W.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank all the participants in the study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejro.2026.100749.

Data availability

Data generated or analyzed during the study are available from the corresponding author by request and upon institutional review.

References

[1] C. Chiyanika, D.F. Chan, S.C. Hui, H. So, M. Deng, D.K. Yeung, et al., The relationship between pancreas steatosis and the risk of metabolic syndrome and insulin resistance in Chinese adolescents with concurrent obesity and non-alcoholic fatty liver disease, Pediatr. Obes. (2020) e12653.

[2] T.T. Chan, Y.K. Tse, R.N. Lui, G.L. Wong, A.M. Chim, A.P. Kong, et al., Fatty pancreas is independently associated with subsequent diabetes mellitus development: a 10-year prospective cohort study, Clin. Gastroenterol. Hepatol. 20 (2022) 2014–2022, e4.

[3] A. Moglia, M. Cavicchioli, L. Mainardi, P. Cerveri, Deep learning for pancreas segmentation on computed tomography: a systematic review: A. Moglia et al, Artif. Intell. Rev. 58 (2025) 220.

[4] S. Kato, A. Iwasaki, Y. Kurita, J. Arimoto, T. Yamamoto, S. Hasegawa, et al., Three-dimensional analysis of pancreatic fat by fat-water magnetic resonance imaging provides detailed characterization of pancreatic steatosis with improved reproducibility, PloS One 14 (2019) e0224921.

[5] X. Li, Q. Yang, H. Ye, S. Li, Y. Wang, W. Yu, Comparison of pancreatic fat content measured by different methods employing MR mDixon sequence, Plos One 16 (2021) e0260001.

[6] J.M. Williams, M.A. Hilmes, B. Archer, A. Dulaney, L. Du, H. Kang, et al., Repeatability and reproducibility of pancreas volume measurements using MRI, Sci. Rep. 10 (2020) 4767.

[7] S. Jain, G. Sikka, R. Dhir, A systematic literature review on pancreas segmentation from traditional to non-supervised techniques in abdominal medical images, Artif. Intell. Rev. 57 (2024) 317.

[8] H.R. Roth, L. Lu, N. Lay, A.P. Harrison, A. Farag, A. Sohn, et al., Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation, Med. Image Anal. 45 (2018) 94–107.

[9] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, et al. Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv: 2102.04306. (2021).

[10] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, et al., Unetr: Transformers for 3d medical image segmentation, (2022) 574–584.

[11] M.E. Rinella, J.V. Lazarus, V. Ratziu, S.M. Francque, A.J. Sanyal, F. Kanwal, et al., A multisociety Delphi consensus statement on new fatty liver disease nomenclature, Hepatology 78 (2023) 1966–1986.

[12] W. Lee, W. Wang, Bariatric surgery: Asia-pacific perspective, Obes. Surg. 15 (2005) 751–757.

[13] H. Yu, A. Shimakawa, C.A. McKenzie, E. Brodsky, J.H. Brittain, S.B. Reeder, Multiecho water-fat separation and simultaneous R estimation with multifrequency fat spectrum modeling, Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med. 60 (2008) 1122–1134.

[14] H. Sohrabi. Re: What is the minimum value of correlation coefficient to prove the existence of the accepted relationship between scores of two of more tests, (2015).

[15] S. Kato, A. Iwasaki, Y. Kurita, J. Arimoto, T. Yamamoto, S. Hasegawa, et al., Three-dimensional analysis of pancreatic fat by fat-water magnetic resonance imaging provides detailed characterization of pancreatic steatosis with improved reproducibility, PloS One 14 (2019) e0224921.

[16] A.T. Bagur, P. Aljabar, G.R. Ridgway, M. Brady, D. Bulte, Pancreas MRI segmentation into head, body, and tail enables regional quantitative analysis of heterogeneous disease, medRxiv (2021).

[17] T. Joshi, J. Virostko, M.S. Petrov, Artificial intelligence-based models for quantification of intra-pancreatic fat deposition and their clinical relevance: a systematic review of imaging studies, Eur. Radio. 36 (2026) 627–641.

[18] L.P. Janssens, H. Takahashi, H. Nagayama, F. Nugen, W.R. Bamlet, A.L. Oberg, et al., Artificial intelligence assisted whole organ pancreatic fat estimation on magnetic resonance imaging and correlation with pancreas attenuation on computed tomography, Pancreatology 23 (2023) 556–562.

[19] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, et al., TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers, Med. Image Anal. 97 (2024) 103280.

[20] J.C. Wood, N. Ghugre, Magnetic resonance imaging assessment of excess iron in thalassemia, sickle cell disease and other iron overload diseases, Hemoglobin 32 (2008) 85–96.

[21] C.H. Ho, L. Xiao, K.Y. Kwok, S. Yang, B. Fung, K. Yu, et al., Common artifacts in magnetic resonance imaging: a pictorial essay, Hong. Kong J. Radiol. HKJR Xianggang Fang. She Ke Yi Xue Za Zhi. 26 (2023) 58–65.