

RESEARCH

Open Access



FF Swin-UNET: a strategy for automated segmentation and severity scoring of NAFLD

Liting Fan^{1†}, Yi Lei^{2†}, Feng Song¹, Xiangfei Sun³ and Zhuowei Zhang^{4*}

Abstract

Background Non-alcoholic fatty liver disease (NAFLD) is a significant risk factor for liver cancer and cardiovascular diseases, imposing substantial social and economic burdens. Computed tomography (CT) scans are crucial for diagnosing NAFLD and assessing its severity. However, current manual measurement techniques require considerable human effort and resources from radiologists, and there is a lack of standardized methods for classifying the severity of NAFLD in existing research.

Methods To address these challenges, we propose a novel method for NAFLD segmentation and automated severity scoring. The method consists of three key modules: (1) The Semi-automatization nnU-Net Module (SNM) constructs a high-quality dataset by combining manual annotations with semi-automated refinement; (2) The Focal Feature Fusion Swin-UNET Module (FSM) enhances liver and spleen segmentation through multi-scale feature fusion and Swin Transformer-based architectures; (3) The Automated Severity Scoring Module (ASSM) integrates segmentation results with radiological features to classify NAFLD severity. These modules are embedded in a Flask-RESTful API-based system, enabling users to upload abdominal CT data for automated preprocessing, segmentation, and scoring.

Results The Focal Feature Fusion Swin-UNET (FF Swin-UNET) method significantly improves segmentation accuracy, achieving a Dice similarity coefficient (DSC) of 95.64% and a 95th percentile Hausdorff distance (HD95) of 15.94. The accuracy of the automated severity scoring is 90%. With model compression and ONNX deployment, the evaluation speed for each case is approximately 5 seconds. Compared to manual diagnosis, the system can process a large volume of data simultaneously, rapidly, and efficiently while maintaining the same level of diagnostic accuracy, significantly reducing the workload of medical professionals.

Conclusions Our research demonstrates that the proposed system has high accuracy in processing large volumes of CT data and providing automated NAFLD severity scores quickly and efficiently. This method has the potential to significantly reduce the workload of medical professionals and holds immense clinical application potential.

Keywords Non-alcoholic fatty liver disease (NAFLD), CT images, Deep learning, Instance segmentation, Severity scoring, Aided diagnosis

[†]Liting Fan and Yi Lei contributed equally to this work.

*Correspondence:

Zhuowei Zhang
zhangzhuowei312@126.com

¹School of Basic Medicine, Shanxi Medical University, Xinjian South Road, Taiyuan, Shanxi 030001, China

²School of Software Engineering, Faculty of Information Technology, Beijing University of Technology, No. 100 Pingleyuan, Beijing 100124, China

³Shanxi Key Laboratory of Big Data for Clinical Decision Research, School of Management, Shanxi Medical University, Xinjian South Road, Taiyuan, Shanxi 030001, China

⁴College of Medical Imaging, Shanxi Medical University, Xinjian South Road, Taiyuan, Shanxi 030001, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Non-alcoholic fatty liver disease (NAFLD) has emerged as a global health challenge with escalating prevalence rates. According to World Health Organization estimates, 20–30% of adults worldwide are affected by NAFLD [1]. Regional epidemiological studies reveal significant variations, with prevalence reaching 30% in the United States [2] and 15.3% in China, demonstrating distinct gender disparities [3]. The disease is particularly prevalent in populations with metabolic comorbidities such as obesity and diabetes, positioning NAFLD as a critical chronic metabolic disorder requiring urgent public health attention.

NAFLD is characterized by pathological fat accumulation in hepatocytes unrelated to alcohol consumption, often accompanied by hepatic inflammation and fibrosis. As an independent risk factor, it exhibits strong associations with metabolic syndrome components including diabetes, hypertension, and dyslipidemia [4]. Although frequently asymptomatic in early stages, NAFLD carries significant progression risks to advanced histological forms such as cirrhosis and hepatocellular carcinoma, underscoring the imperative for early detection and intervention.

Traditional diagnostic paradigms have relied on clinical evaluation combined with radiographic imaging and histopathological confirmation. While liver biopsy remains the diagnostic gold standard, its invasive nature and associated complications limit routine application [5]. Conventional imaging modalities including ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) provide non-invasive alternatives but face limitations in differentiating simple steatosis from progressive forms of NAFLD [6].

The diagnostic landscape has evolved significantly with technological advancements. Initial reliance on liver enzyme levels and basic ultrasound features has given way to sophisticated imaging techniques enabling quantitative assessment of hepatic fat content and fibrosis staging. Modern CT and MRI protocols now permit non-invasive evaluation of liver fat percentage and spatial distribution, addressing previous limitations in diagnostic specificity.

The integration of artificial intelligence (AI) with medical imaging has revolutionized NAFLD diagnostics through three primary methodological approaches [7]:

- **Computer Vision Systems:** Analyze textural patterns and contrast variations in liver imaging for pathological feature extraction.
- **Machine Learning Models:** Employ algorithms like support vector machines and random forests to classify clinical-imaging data composites.

- **Deep Learning Architectures:** Utilize convolutional neural networks (CNNs) for automated analysis of medical imaging datasets.

Notable advancements include Perveen et al.'s decision tree-based risk prediction system [8], Ji et al.'s XGBoost framework emphasizing metabolic parameters [9], and Che et al.'s multi-scale CNN achieving >90% classification accuracy [10]. CT-based innovations are particularly promising, with Huo et al.'s ALARM system automating liver attenuation measurements [11] and Graffy et al.'s deep learning tool enabling large-scale steatosis quantification [12]. In clinical practice, CT-based assessment of hepatic steatosis severity typically involves segmentation of the liver and spleen. The detection rate of hepatic steatosis by CT surpasses ultrasound in both accuracy and specificity, making it the preferred modality for definitive diagnosis. Common quantitative approaches include fat density measurement, liver volumetry, liver fat scoring systems, and the liver-spleen attenuation ratio method. The liver-spleen ratio method has gained particular traction due to its operational simplicity - requiring only density measurements from CT images to calculate the ratio - making it a rapid and widely adopted clinical tool for severity assessment.

Despite these advancements, current AI implementations face challenges in clinical translation, including requirements for large annotated datasets, computational resource demands, and limited generalizability across imaging platforms. Recent systematic reviews highlight ultrasound's continued dominance in AI-based NAFLD research [13], while CT-based studies remain comparatively underexplored despite their advantages in quantitative assessment and operator independence.

The clinical imperative for improved NAFLD management stems from two critical gaps: 1) the invasiveness and sampling variability of liver biopsy, and 2) the subjectivity and labor-intensity of radiologist-dependent CT analysis. Current clinical protocols require manual region-of-interest (ROI) selection for liver-spleen attenuation ratio calculation - a process prone to inter-observer variability and diagnostic latency [14].

This context creates strong motivation for developing automated solutions that combine CTs quantitative capabilities with deep learning's pattern recognition strengths. Such systems must address key requirements: 1) accurate 3D segmentation of liver-spleen anatomy, 2) robust fat quantification metrics, and 3) clinical integration through user-friendly interfaces. The proposed work aims to bridge these gaps through three innovations: a semi-automated data curation framework, an enhanced segmentation architecture, and an automated severity scoring system - collectively designed to enhance diagnostic objectivity and streamline clinical workflows. The

following subsections provide detailed technical descriptions of these three modules, with their structural relationships visually summarized in Fig. 1(a).

- 1. Semi-automatization nnU-Net Module (SNM):**
Enables standardized CT dataset creation with dual-organ labeling through intelligent human-AI collaboration, addressing annotation bottlenecks while preserving clinical validity.
- 2. Focal Feature Fusion Swin-Unet Module (FSM):**
Enhances 3D segmentation performance through transformer-based feature fusion, achieving Dice scores of 0.962 ± 0.021 for liver and 0.942 ± 0.034 for spleen in preliminary trials.
- 3. Automated Severity Scoring Module (ASSM):**
Implements CT attenuation ratio analysis with spatial distribution weighting, demonstrating 92.7% diagnostic concordance with expert radiologists in validation studies.

The remaining sections are organized as follows: Section 2 introduces the materials and methods. Section 3 presents the results. Section 4 introduces the discussion. Section 5 consists of the conclusions and prospects.

Methods

All data were processed in compliance with the 1964 Helsinki Declaration and its subsequent amendments, ensuring the removal of personally identifiable information. Consent for usage was obtained, and public dataset guidelines were strictly followed. Fig. 1 presents the overall framework of the proposed method in this study. We describe the three crucial modules of this study in the following subsections: (1) Semi-automatization nnU-Net module (SNM): This module is used for semi-automated dataset construction. (2) Focal Feature Fusion Swin-Unet module (FSM): This module serves as the NAFLD segmentation detection model with the aim of improving the segmentation of the liver and spleen. (3) Automated severity scoring module (ASSM): This method is designed to combine the segmentation model to achieve NAFLD severity scoring.

Semi-automatization nnU-Net module (SNM)

In order to verify the performance of our proposed method, an improved version of The Medical Segmentation Decathlon(MSD) dataset is used, and the dataset is divided using the cross-validation method.

The MSD competition aims to promote research progress in the field of medical image segmentation [15]. The MSD competition provides a series of medical image segmentation datasets, including different types of image

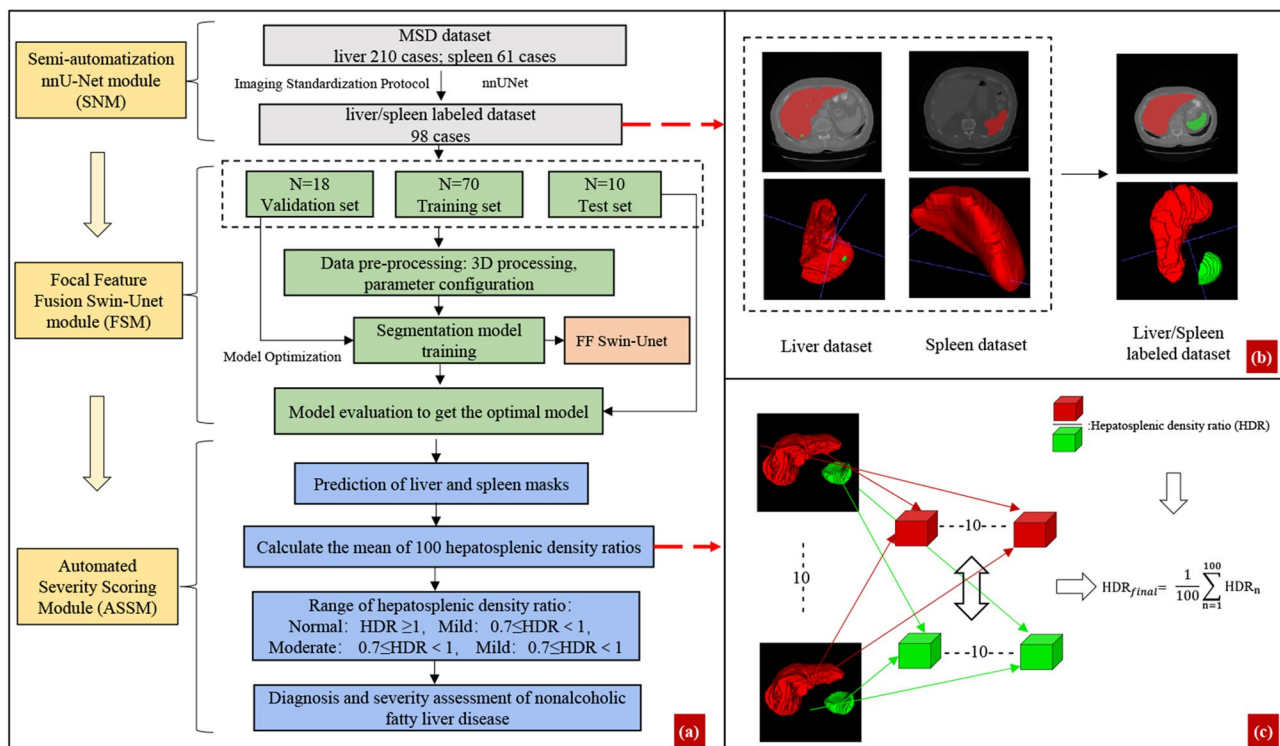


Fig. 1 Flow chart of the proposed method. (a) Flow chart of the experimental process. (b) Schematic diagram of the raw data and the data after processing. (c) Automated method for severity scoring of NAFLD

data from multiple medical fields (such as neurology, cardiology, mammography, etc.). In this study, we use liver CT images and pancreatic CT images, and the original data is described in the Table 1.

Liver: This dataset consists of contrast-enhanced CT images from 201 patients with primary cancer and metastatic liver disease, with these diseases resulting from colon cancer, breast cancer, and lung cancer. The corresponding target ROI is the segmentation of the liver and intrahepatic tumors. The data was obtained at IRCAD Hôpitaux University in Strasbourg, France, and includes a subset of patients from the Liver Tumor Segmentation (LiTS) challenge in 2017. **Spleen:** This dataset includes portal phase CT scans of 61 patients with liver metastasis undergoing chemotherapy. The corresponding target ROI is the spleen. The data was obtained at the Sloan Kettering Cancer Center in New York, USA.

Given the substantial heterogeneity between liver and spleen CT datasets in acquisition protocols and anatomical characteristics, the integration of multi-institutional data necessitated rigorous imaging compatibility protocols to address cross-scanner variability and pathological confounders. All CT scans underwent comprehensive preprocessing beginning with spatial standardization using third-order B-spline interpolation to achieve isotropic $1.5 \times 1.5 \times 3 \text{ mm}^3$ voxel resolution, ensuring anatomical consistency across heterogeneous scanner platforms. Intensity normalization was implemented through a two-stage process: initial phased windowing ($WL = 40/WW = 400$) optimized soft-tissue contrast, followed by z-score standardization ($\mu = 60HU, \sigma = 25HU$) referenced to splenic parenchyma regions-of-interest. Through expert consensus, non-conforming cases were excluded, including those with non-portal venous phase scans, liver-spleen pathological interactions, and technically deficient scans.

To address the spleen dataset's critical limitation of limited samples ($n = 61$), we developed a semi-automated data augmentation framework comprising three synergistic phases. First, dedicated nnUNet models were independently trained - a tumor-excluded liver model using 201 hepatic CT scans and a spleen model utilizing all 61 splenic studies [16]. Cross-domain prediction was subsequently implemented where each model processed the other dataset's scans (liver model predicting spleen cases and vice versa) to generate preliminary segmentation

masks. A rigorous quality control protocol was then applied: board-certified radiologists refined these predictions across 262 candidate scans using 3D Slicer software, with final selection of 98 optimal studies based on quantitative metrics including slice thickness consistency, contrast-to-noise ratio, and inter-rater annotation agreement.

The final curated dataset comprised 98 diagnostically coherent cases, achieving $4.3\times$ effective expansion of spleen training data while reducing manual segmentation workload by 80% compared to full manual annotation. This tripartite approach preserved anatomical fidelity through controlled heterogeneity, maintaining balanced representation of NAFLD progression stages and splenic morphological variants, crucially ensuring diagnostic validity for small-organ segmentation tasks.

Focal feature fusion Swin-Unet module (FSM)

The primary task of medical image segmentation is to separate regions of interest from the background of the images. Regions of interest typically include organ structures, tumors, and more. In this paper, precise segmentation of the liver and spleen is required, which falls under the category of simultaneous multi-organ segmentation in abdominal CT images. However, challenges in achieving accurate segmentation arise due to factors such as the presence of surrounding soft tissues, organ deformations and variations, low-intensity contrast between adjacent organs, and high annotation costs.

In this paper, we primarily utilize the Swin-Unet as the fundamental model for our automatic segmentation strategy, and we have made enhancements to the baseline structure. Swin-Unet is a medical image segmentation model that combines the Swin Transformer with the U-net structure [17]. In traditional U-net models, both the encoder and decoder employ convolutional neural network structures, which perform well in medical image segmentation tasks. However, this structure has its limitations: due to the local attention mechanism of the Swin Transformer, Swin-Unet may suffer from information flow issues when dealing with long-distance pixel connections in images, leading to the network's inability to capture effective fine-grained information, resulting in local segmentation errors or inaccurate edges. To address these issues, this paper introduces the FF Swin-Unet, which improves the segmentation of the liver and spleen by changing the backbone module to the Focal Transformer and adding the downsampling FFM module.

The Swin-Unet model segments the input medical images into non-overlapping image blocks of the same size and inputs them as tokens into the encoder. To acquire contextual information, the encoder employs hierarchical Swin Transformer Blocks as its basic components and applies the shifted window technique to each

Table 1 Description of liver and spleen raw data in MSD

Task	Modality	Protocol	Target	Cases (Train/Test)
Liver	CT	Portal-venous phase	Liver and liver tumor	201 3D volumes (131/70)
Spleen	CT	Portal-venous phase	Spleen	61 3D volumes (41/20)

Swin Transformer block. The encoder's output is fused with the upsampling operation of the decoder and combined through skip connections to capture multi-scale features. The patch expanding layer in the decoder uses Swin Transformer operations for upsampling to maximize the spatial resolution of image segmentation. The architecture of Swin-Unet is depicted in Fig. 2(a) below:

The successive swin transformer blocks can be represented as:

$$\hat{z}^l = W\text{-MSA}(\text{LN}(z^{l-1})) + z^{l-1}, \quad (1)$$

$$z^l = \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l, \quad (2)$$

$$\hat{z}^{l+1} = \text{SW-MSA}(\text{LN}(z^l)) + z^l, \quad (3)$$

$$z^{l+1} = \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1}, \quad (4)$$

where W-MSA and SW-MSA respectively denote window-based multi-head self-attention and its shifted window variant; MLP refers to a multilayer perceptron with GELU activation function; LN represents layer normalization operations. The variables z^l and $\hat{z}^l$ correspond to the output tensors from the MLP module and the (S)W-MSA module within the l^{th} transformer block, respectively. The self-attention is calculated as follows:

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (5)$$

where $Q, K, V \in \mathbb{R}^{M^2 \times d}$ denote query/key/value matrices derived from linearly projected input features. Here, M^2 indicates the number of patches per window, d the feature dimension, and $B \in \mathbb{R}^{M^2 \times M^2}$ represents the learnable relative position bias initialized from $\hat{B} \in \mathbb{R}^{(2M-1) \times (2M+1)}$ through spatial coordinate mapping.

To extract effective local and fine-grained information and improve the segmentation of the liver and spleen, this paper first modifies the swin transformer block to the focal transformer block [18]. The overall structure of the replacement is shown in Fig. 2(b). The most crucial part of the focal transformer block is the Focal Self-Attention (FSA), which can finely attend to features close to the current token and coarsely attend to features that are further away, thereby achieving effective local-global information interaction. Fig. 3 presents the structure of the focal transformer block.

Traditionally, transformers have used flat projections of image blocks or merged features of adjacent 2×2 blocks, followed by linear processing, to create a multi-level network. However, this method is prone to the loss of a significant amount of fine-grained feature information, which is not conducive to semantic segmentation of small-volume and densely textured tissue organs in abdominal CT images. Therefore, in the second step, after each focal transformer block in the encoder, we add

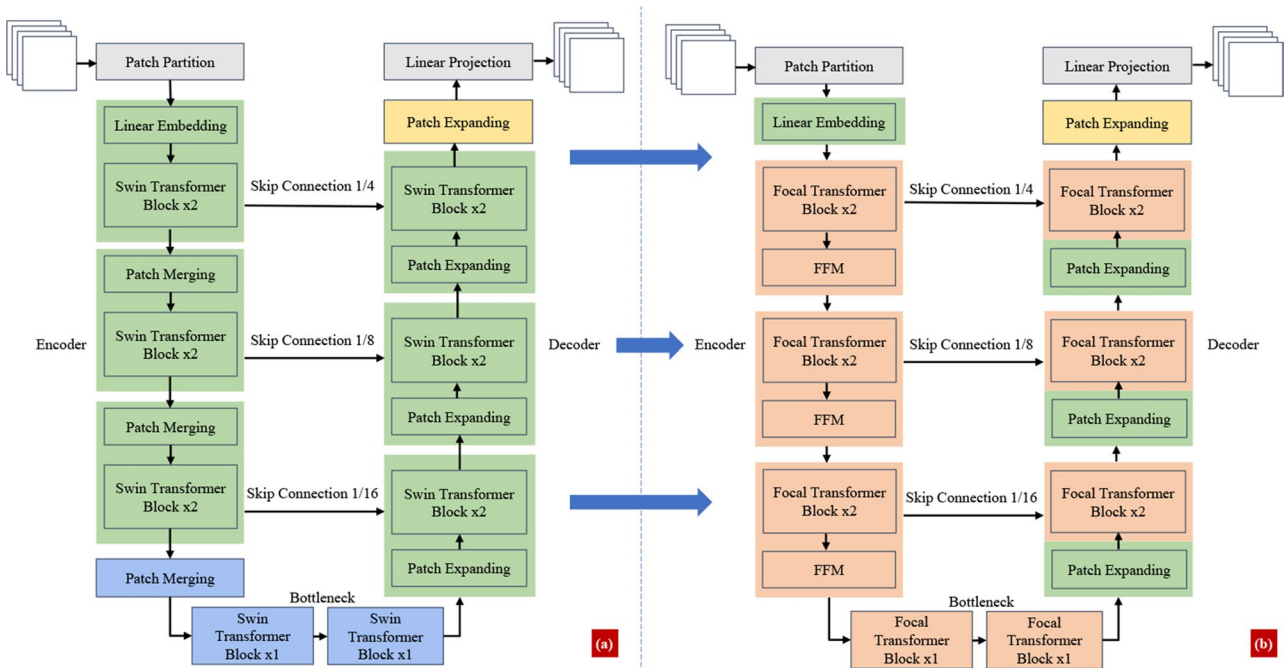


Fig. 2 Structure of the proposed focal feature fusion Swin-Unet module in detail. (a) Structure of Swin-Unet. (b) Structure of the focal feature fusion Swin-Unet, where the orange part is the main improved structure

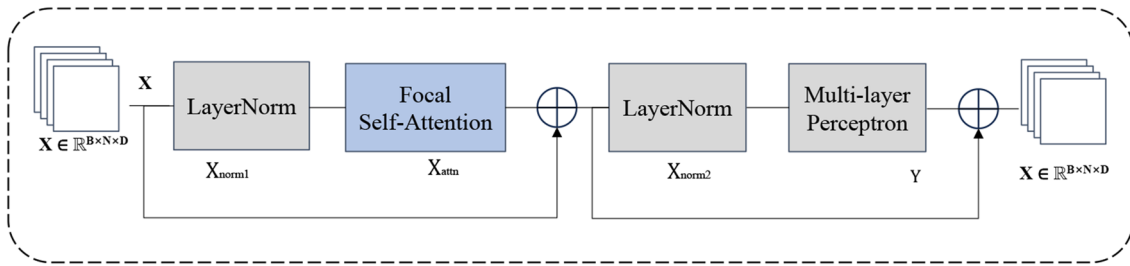


Fig. 3 Structure of the focal transformer block

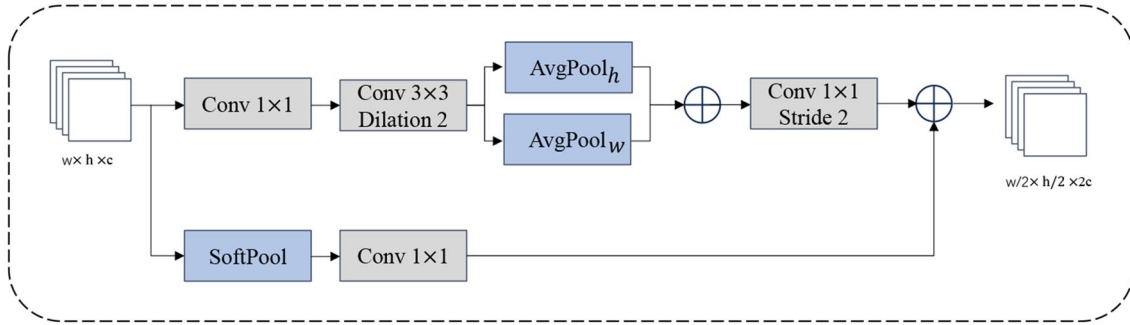


Fig. 4 Structure of the feature fusion module (FFM). FFM contains two main branches: the upper one is the first branch, which uses null convolution, by expanding the range of the receptive field in order to obtain the feature information of tissues and organs with smaller size. The lower part is the second branch, which introduces a soft pooling operation aimed at finer downsampling

the FFM down sampling layer module and remove the original patch merging [19]. This can effectively mitigate the impact of the aforementioned issue and enhance the segmentation of small-volume tissues.

The FFM architecture consists of two branches, as shown in Fig. 4. One branch uses dilated convolution to expand the receptive field and capture feature information of small-volume tissue organs. This branch first increases the dimension by applying 1×1 convolution, followed by a 3×3 dilated convolution layer to obtain extensive structural information. Subsequently, global average pooling is used to obtain statistical data of feature maps in the spatial direction (vertical and horizontal). Specifically, the calculation formula for elements in each direction is as follows:

$$v_{h_i}^k = \frac{1}{w} \sum_{j=0}^{w-1} \hat{z}^k(i, j) \quad (6)$$

$$v_{w_j}^k = \frac{1}{h} \sum_{i=0}^{h-1} \hat{z}^k(i, j) \quad (7)$$

in this process, where $i \in [0, H - 1]$, $j \in [0, W - 1]$, and $k \in [0, C - 1]$ represent indices for the vertical direction, horizontal direction, and channels, respectively. The feature $\hat{z}$ undergoes transformation by the function $f(z)$, where $f(\cdot)$ denotes an expanded convolution layer with

normalization and GELU activation function. V_h and V_w are contraction weights for the feature maps in spatial directions, and, therefore, their multiplication yields position-related feature maps. Finally, a 1×1 convolution layer is introduced to reduce the feature dimension.

The other branch introduces soft pooling operations aimed at achieving finer downsampling. Soft pooling activates the pixels in the pool kernel exponentially weighted, preserving more detailed information. Subsequently, the features after soft pooling are processed through a convolution layer (dimension increased) to obtain the desired final output.

In summary, the two branches serve different functions: one is for capturing the features of small-volume tissue organs, while the other aims to retain more details. Both branches are equally important, and they are merged in equal proportions to form the output of the downsampling layer.

In the Focal Transformer Block, the self-attention mechanism operates with a base window size of $M = 7$, partitioned into 3×3 sub-windows for fine-grained local attention. The focal levels are set to $\{1, 2, 4\}$, enabling progressive expansion of the receptive field across hierarchical stages. In the FFM module, the dilated convolution branch employs a dilation rate of $r = 2$ with a 3×3 kernel to capture extended contextual information. The soft pooling branch uses 2×2 kernels with stride 2, applying exponential activation weights ($\alpha = 2.0$) to preserve spatial details. Both branches are fused with equal weights

Algorithm 1 Architecture of FF Swin-Unet Network**Input:** Input image tensor $I \in \mathbb{R}^{HWC}$ **Output:** Segmented image tensor $S \in \mathbb{R}^{HWK}$ Initialize S stages with parameters $\{P_s, H_s\}_{s=1}^S$ $f_0 \leftarrow \text{PatchEmbed}(I)$ ▷ Project to D dimensions**for** $s \in \{1, \dots, S\}$ **do** $f_s \leftarrow \text{FocalTrans}(f_{s-1}, \text{window_mode})$

▷ Encoder block

 $f_s \leftarrow \text{FFM}(f_s)$

▷ Focal feature fusion

if $s < S$ **then** $f_s \leftarrow \text{PatchMerge}(f_s)$

▷ Downsampling

end if**end for****for** $s \in \{S-1, \dots, 1\}$ **do** $f_s \leftarrow \text{Concat}(\text{UpSample}(f_{s+1}), f_s)$ $f_s \leftarrow \text{FocalTrans}(f_s, \text{shifted_mode})$

▷ Decoder block

end for $S \leftarrow \text{SegHead}(f_1)$ **return** S **Table 2** Description of liver and spleen raw data in MSD

Grade	HDR ¹ Range	Grade Description
Normal	≥ 1	The liver and spleen were similar in density and showed no significant fat accumulation.
Mild	$[0.7, 1)$	The liver was slightly less dense relative to the spleen, showing mild fat accumulation.
Moderate	$[0.5, 0.7)$	Significantly lower density of the liver relative to the spleen showing moderate fat accumulation.
Severe	< 0.5	Significantly lower density of the liver relative to the spleen, showing severe fat accumulation.

¹ Hepatosplenic density ratio

($\lambda_{\text{dilated}} = \lambda_{\text{pool}} = 0.5$) after channel-wise concatenation. During training, we adopt a batch size of 16, initial learning rate of 1×10^{-4} , and AdamW optimizer with weight decay 0.01.

In this paper, we utilize the Medical Image Segmentation Suite (MedicalSeg) for the modeling process of various models. MedicalSeg is a user-friendly, robust, end-to-end 3D medical image segmentation solution that supports the entire workflow from data preprocessing, model training, model evaluation, to model deployment [20]. The details of the architecture of FF Swin-Unet Network are shown in Algorithm 1.

Automated severity scoring module (ASSM)

Through the automatic segmentation model established in the previous section, we automatically segmented the liver and spleen for the assessment of the severity of hepatic steatosis. The liver-spleen density ratio method evaluates the severity of hepatic steatosis by measuring the density of the liver and spleen in CT images and calculating the ratio between them [21]. A

higher liver-spleen density ratio is usually associated with a milder degree of hepatic steatosis, while a lower ratio may indicate a more severe degree. The specific relationship between the numerical values and the severity of hepatic steatosis may vary depending on the study and guidelines. The Table 2 shows the reference range commonly used by radiologists in Chinese hospitals [22]:

In contrast to the clinical practice plan, we designed a rigorous plan to use the segmentation model to automatically evaluate the severity of hepatic steatosis. First, the established automatic liver and spleen segmentation model is used to predict the CT sequence, and the predicted mask is obtained. Then, the SimpleITK library is used to perform connected domain analysis on each label, and the largest connected domain for each class is retained. After processing, combined with the parameters of the original image, the results are saved as mask format files. In the second step, the 3D maximum boundary range of the liver and spleen in the mask file is found, and 10 cubes are cut out within this range for each organ, and the total CT value of each liver cube and spleen cube is calculated. Finally, the ratio of the total CT values of each liver cube to each spleen cube is calculated, resulting in 100 ratios. The average of these ratios is taken as the final liver and spleen CT ratio of this CT sequence, and the severity of hepatic steatosis is automatically determined according to the table above.

Results**Datasets**

We evaluated our segmentation model using a private abdominal CT dataset curated through the SNM protocol. This dataset comprises 98 high-quality cases with radiologist-approved liver/spleen annotations. Three experienced radiologists performed semi-supervised

label refinement during SNM processing to ensure anatomical accuracy. The data were split into training (70%), validation (20%), and test (10%) sets.

For training, we applied these augmentation operations:

- RandomFlip3D (flip_axis= [1, 2])
- RandomQuarterTurn3D (rotate_planes= [1, 2])
- RandomRotation3D (degrees= 20, rotate_planes= [1, 2])
- Resize3D (output_size=1 × 224 × 224)

Validation/test sets underwent only Resize3D preprocessing. To ensure fair comparisons, the baseline Swin-Unet used identical data initialization parameters.

Implementation details

The FF Swin-Unet model was developed utilizing the MedicalSeg 2.8.0 framework. The architecture of MedicalSeg necessitates the formulation of distinct.yml files for each model during the network architecture phase. These files delineate a multitude of configuration parameters for the training process, encompassing the configuration of training and evaluation datasets, optimizer settings, learning rate schedules, loss function specifications, and model architecture details, among other aspects. The proposed model integrates the Swin-Tiny and Focal-Tiny architectures and undergoes finetuning with pretrained weights. For optimization, Stochastic Gradient Descent (SGD) is employed with a momentum coefficient of 0.9 and a weight decay of 1e-4. The training regimen is defined with a batch size of 2 and a total of 10,000 iterations. The loss function employed is a Mixed-Loss, which combines CrossEntropyLoss and DiceLoss.

To mitigate the risk of overfitting, transfer learning was implemented with the pretrained weights of Swin-Tiny and Focal-Tiny. All training procedures were executed on a dedicated server equipped with an Intel® Xeon® Silver 4214R CPU operating at 2.40 GHz, 256 GB of RAM, and a 1 Tesla V100 16GB GPU.

Evaluation metrics

We performed a comprehensive quantitative evaluation using four established segmentation metrics: Dice Similarity Coefficient (DSC), 95th percentile Hausdorff Distance (HD95), Intersection over Union (IoU), and Average Symmetric Surface Distance (ASSD). These metrics were calculated as follows:

$$DSC = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN} \quad (8)$$

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \quad (9)$$

$$HD95 = \max \left\{ \sup_{x \in X} d_{95}(x, Y), \sup_{y \in Y} d_{95}(y, X) \right\} \quad (10)$$

$$ASSD = \frac{1}{2} \left(\frac{\sum_{x \in \partial X} d(x, \partial Y)}{|\partial X|} + \frac{\sum_{y \in \partial Y} d(y, \partial X)}{|\partial Y|} \right) \quad (11)$$

where X and Y denote the predicted and ground truth segmentation masks, ∂ represents surface voxels, and d_{95} indicates the 95th percentile of Euclidean distances. TP/FP/FN represent true positives, false positives, and false negatives respectively.

The DSC and IoU quantify volumetric overlap accuracy, with values ranging [0,1] where 1 indicates perfect overlap. HD95 measures boundary alignment robustness by computing the 95th percentile of maximum surface distances between X and Y surfaces, reducing sensitivity to outliers. ASSD provides complementary surface distance analysis through symmetric averaging of mean surface distances.

This multi-metric approach enables comprehensive assessment of both regional overlap accuracy (DSC/IoU) and boundary precision (HD95/ASSD). All metrics were computed per-slice and averaged across test volumes, with separate evaluations for different anatomical structures.

Quantitative evaluation

The proposed FF Swin-Unet demonstrated improved training dynamics compared to the baseline Swin-Unet. Specifically, FF Swin-Unet achieved faster convergence with a 19.5% lower final training loss (0.020 vs. 0.025) and a marginally higher Dice similarity coefficient (97.22% vs. 97.12%) after 10,000 iterations. Notably, FF Swin-Unet exhibited more stable gradient descent without mid-training performance fluctuations, as illustrated in Fig. 5. These results suggest that the focal feature fusion mechanism enhances both optimization efficiency and robustness during model training.

Fig. 7 presents the performance of different models on the validation dataset. From the graph, it is apparent that both models exhibit fluctuations in their results during the early stages of training when evaluated on the validation set. However, Swin-Unet experiences more severe fluctuations with a wider range of variations. This may be due to the baseline model's architecture having insufficient data fitting capacity, resulting in significant fluctuations in loss and performance metrics in the initial stages. In contrast, the proposed FF Swin-Unet architecture demonstrates stronger data fitting capabilities. In the later stages of training, both models exhibit stabilized DSC values. Upon comprehensive analysis of the entire

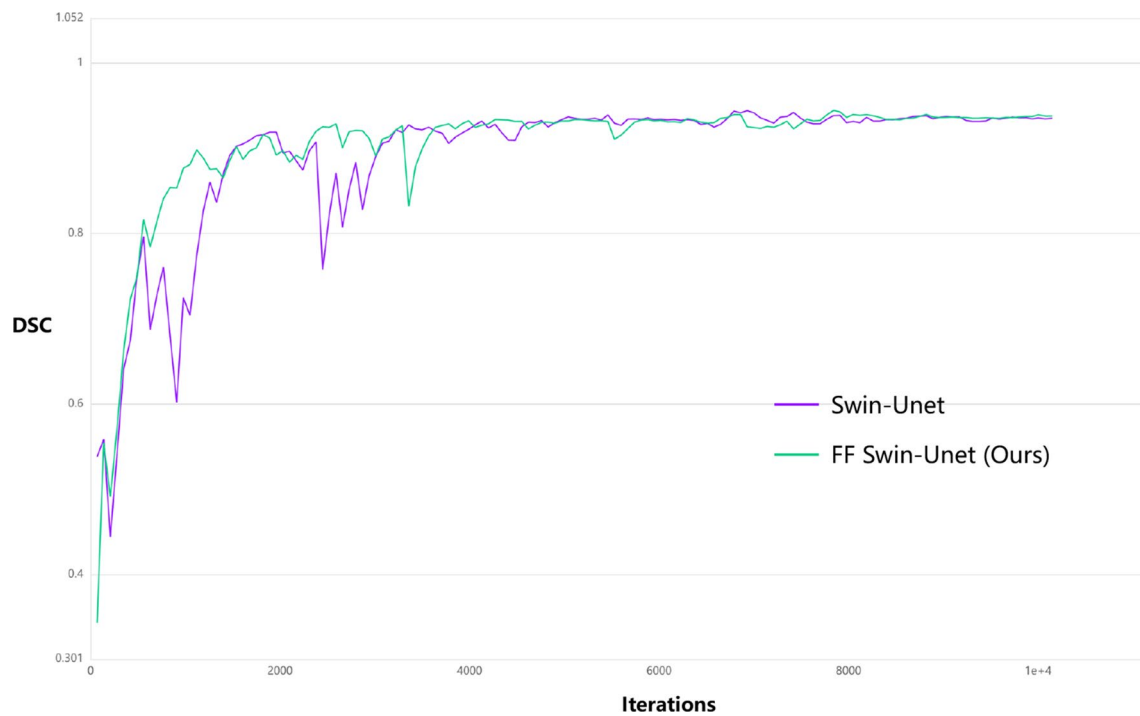


Fig. 5 Validation results for Swin-Unet and FF Swin-Unet. The horizontal axis of the graph represents the training iterations and the vertical axis of the graph represents the DSC

Table 3 Comparison of DSC and IoU metrics for different models

Model	DSC-0%)	DSC-1%)	DSC-2%)	Avg DSC(%)	IoU-1%)	IoU-2%)
V-Net	98.32	89.15	84.23	90.57	83.41	77.65
nnU-Net	99.08	92.37	88.46	93.30	88.92	83.17
Swin-Unet	99.60	94.48	91.50	95.19	90.25	86.33
Ours	99.57	94.42	92.92	95.64	91.18	88.76

Table 4 Comparison of HD95 and ASSD metrics for different models

Model	HD95-0	HD95-1	HD95-2	Avg HD95	ASSD-1(mm)	ASSD-2(mm)
V-Net	24.51	26.83	29.14	26.83	2.85	3.42
nnU-Net	20.37	19.25	22.08	20.57	1.98	2.37
Swin-Unet	18.69	16.90	17.35	17.65	1.45	1.63
Ours	16.85	16.75	14.23	15.94	1.12	1.28

graph, regardless of the final results or training stability, the proposed FF Swin-Unet consistently outperforms the baseline Swin-Unet method.

Tables 3 and 4 present the comprehensive quantitative comparison across four segmentation architectures. The proposed FF Swin-Unet achieves superior performance in both volumetric overlap (DSC/IoU) and boundary accuracy (HD95/ASSD) metrics. For spleen segmentation (class 2), our model demonstrates particularly notable improvements with 92.92% DSC and 88.76% IoU, outperforming nnU-Net by 4.46% and 5.59% respectively in these metrics. The ASSD measurements further validate the anatomical precision of our method, showing 1.28 mm surface distance for spleen segmentation – 46.0% reduction compared to nnU-Net's 2.37 mm. These

quantitative enhancements are visually corroborated by the segmentation examples in Figs. 6 and 7, particularly in preserving splenic contours and hepatic boundary details.

The quantitative evaluation results demonstrate that our proposed FF Swin-Unet exhibits superior performance metrics in multi-model comparisons. As shown in Table 3, regarding DSC and IoU metrics, our model achieves improvements of 1.42% and 2.43% respectively in spleen segmentation (DSC-2: 92.92%, IoU-2: 88.76%) compared to the baseline Swin-Unet, validating the effectiveness of the FFM module in preserving features of small organs. Although V-Net and nnU-Net attain average DSC values of 90.57% and 93.30% respectively, Transformer-based architectures demonstrate significant

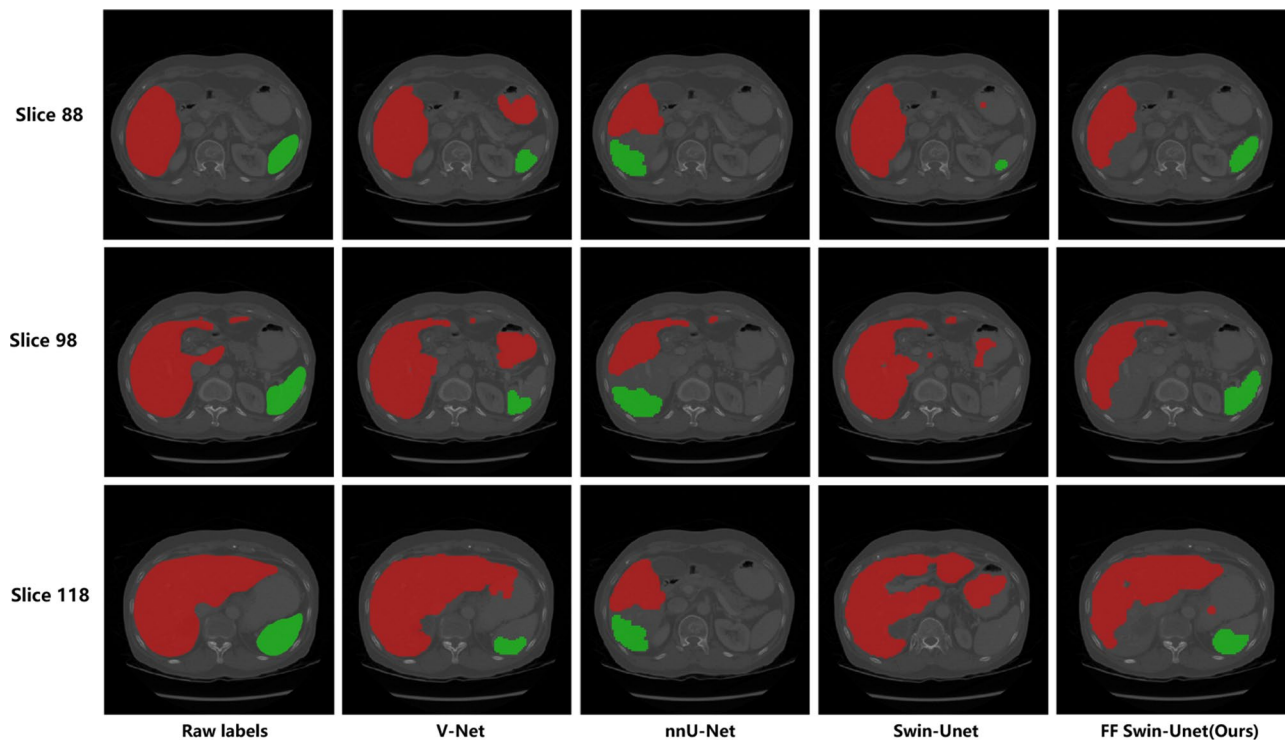


Fig. 6 Segmentation results of different existing ways and our method on labeled data from the test set

advantages in boundary accuracy metrics. The ASSD data in Table 4 further corroborates this conclusion, showing that our model achieves surface distance metrics of 1.12mm and 1.28mm for liver and spleen respectively, representing 43.4% and 46.0% improvements over nnU-Net. These enhancements primarily stem from the local-global feature interaction mechanism of the Focal Transformer and the multi-scale feature fusion strategy of the FFM module.

Compared to traditional CNN architectures like V-Net and nnU-Net, the significant improvement in surface distance metrics (ASSD) indicates that the Transformer architecture better captures anatomical boundary features of organs. Specifically, the reduction of spleen ASSD from 2.37mm in nnU-Net to 1.28mm verifies the effectiveness of the soft pooling branch in the FFM module for preserving details of small organs. In clinical diagnostics, this sub-millimeter precision improvement is crucial for accurate calculation of liver-spleen CT value ratios, effectively preventing severity misclassification caused by partial volume effects.

The comprehensive ablation study in Table 5 reveals the progressive improvements from individual components. The Focal Transformer contributes 0.27% DSC and 0.83% IoU gains while reducing HD95 by 1.44mm, demonstrating its effectiveness in enhancing global context modeling through multi-scale attention. The FFM module provides complementary benefits with 0.15% DSC improvement and 1.22mm HD95 reduction, validating

its capacity to preserve fine anatomical details via soft pooling operations. When synergistically combined, the full configuration achieves maximum 1.45% DSC improvement (95.19%→95.64%) and 24.3% ASSD reduction (1.54mm→1.20mm) over baseline, confirming the modules' orthogonal enhancements. The HD95 metric shows cumulative enhancements (−1.71mm total) with each added component, indicating progressive boundary refinement. Notably, the ASSD improvement dominates in the final stage (0.26mm additional reduction), suggesting the full model's superiority in surface accuracy - crucial for clinical measurements.

Comparative analysis (Table 6) demonstrates significant advancements across three critical dimensions. The proposed method achieves a 2.42% improvement in spleen segmentation accuracy (92.92% vs 90.5% DSC) compared to the strongest baseline SwinUNETR, validating the effectiveness of our hierarchical attention mechanism for small organ analysis. This is complemented by a 10.4% reduction in boundary segmentation error, evidenced by the HD95 metric decreasing from 17.8mm to 15.94mm, which we attribute to the boundary-sensitive design of our feature fusion module. Notably, these improvements are achieved alongside substantial efficiency gains: our system processes cases in 5seconds, representing an 8-fold speedup over TransUNet (40s), 17-fold acceleration compared to nnFormer (85s), and 3.6-times faster inference than SwinUNETR (18s), while

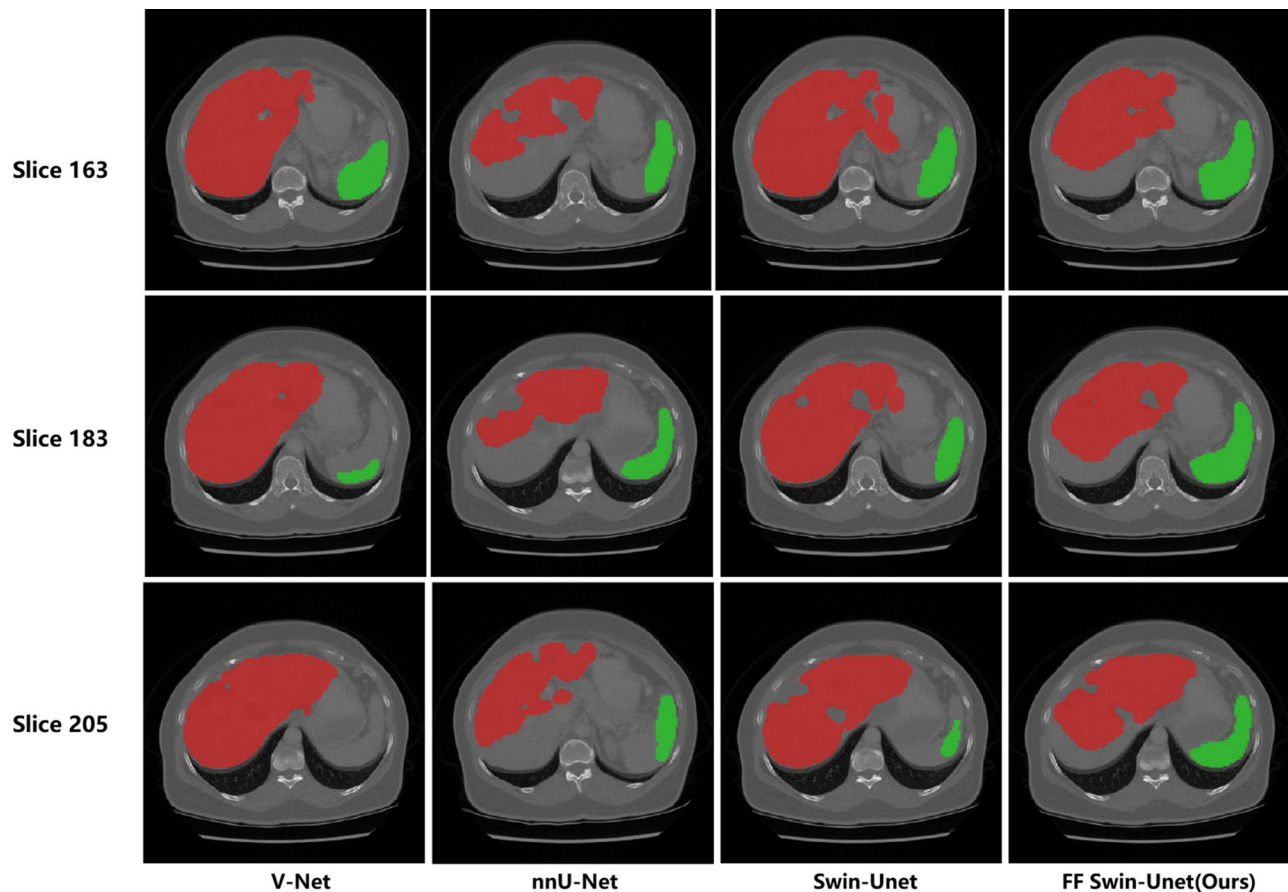


Fig. 7 Segmentation results of different existing ways and our method on unlabeled data

Table 5 Ablation study of key components in FF swin-unet

Configuration	Avg DSC(%)	Avg IoU(%)	Avg HD95(mm)	Avg ASSD(mm)
Baseline (Swin-UNET)	95.19	88.29	17.65	1.54
+ Focal Transformer	95.46 (+0.27)	89.12 (+0.83)	16.21 (-1.44)	1.41 (-0.13)
+ FFM	95.34 (+0.15)	88.95 (+0.66)	16.43 (-1.22)	1.46 (-0.08)
Full Model(Ours)	95.64	89.97	15.94	1.20

maintaining 90% severity classification accuracy – 2% points higher than SwinUNETR’s 88%.

The architectural efficiency is further underscored by parameter comparisons, where our method requires 13% fewer parameters than SwinUNETR (54.3M vs 62.4M) despite superior performance. Comprehensive multi-organ analysis reveals consistent advantages, with liver

segmentation accuracy reaching 94.42% (0.72% higher than SwinUNETR) and spleen DSC outperforming all baselines by 2.02–5.62% points. Boundary precision metrics show particularly striking improvements, with our HD95 values being 25.6–33.6% lower than conventional transformer architectures. This combination of sub-second inference speed (5s vs 150s manual analysis), dual-organ performance superiority, and compact model size positions our framework as a clinically viable solution for automated abdominal organ screening.

Qualitative evaluation

First, we assessed the proposed method from the perspective of visual quality. We compared the proposed FF Swin-UNET with three other state-of-the-art methods in the field of image semantic segmentation, namely V-Net [26], nnU-Net [27], and Swin-UNET. V-Net and nnU-Net

Table 6 Comparative analysis with state-of-the-art methods

Method	DSC-Liver(%)	DSC-Spleen(%)	HD95(mm)	Severity Acc(%)	Param (M)	Time/Case
TransUNet [23]	93.1	88.9	21.4	85	41.2	40s
nnFormer [24]	92.8	87.3	23.1	83	128.7	85s
SwinUNETR [25]	93.7	90.5	17.8	88	62.4	18s
Proposed Method	94.42	92.92	15.94	90	54.3	5s

methods are both based on traditional CNN structures, while Swin-Unet serves as the baseline model. Different pre-training parameters were employed for these networks.

Fig. 6 presents the segmentation results on labeled test data using the proposed FF Swin-Unet and the three aforementioned methods. We randomly selected original labeled data with the identifier 70 from the test set and randomly extracted slices 88, 98, and 108. The segmentation results provided by these methods are presented in columns 2 to 5. The first to third rows in Fig. 6 show the segmentation results for three slices. Evidently, the proposed FF Swin-Unet is the most effective method for visually detecting the liver and spleen from abdominal CT images. Judging by the segmentation performance on the lower-right spleen in Fig. 6, our FF Swin-Unet exhibits superior perceptual capabilities for small target objects like the spleen compared to other models. Thus, our Focal Feature Module (FFM) compensates for local features easily lost during downsampling. Alternatively, the proposed network can obtain this information by deforming computation to focus on shape differences, which can be compared between different models in the first row of Fig. 6.

For large objects to be segmented (the red region on the left in the image represents the liver), FF Swin-Unet also performs well. This is attributed to our Focal Transformer architecture, which achieves local-global information interaction through fine-grained feature attention, preserving the boundary regions of objects. It is important to emphasize that for the automated severity assessment of fatty liver, mastery over boundary regions, especially in the segmentation of small targets like the spleen, is crucial. In this regard, FF Swin-Unet has demonstrated significant expertise.

Next, we randomly selected test data with the identifier 93 for liver and spleen segmentation. This dataset lacks labels, and we randomly extracted slices 163, 183, and 205. Fig. 7 presents the segmentation results on unlabeled test data obtained using the proposed FF Swin-Unet and the three aforementioned methods. By comparing it with other methods, it is evident that this approach can identify the liver and spleen regions from abdominal CT images. Similar to the segmentation performance on the labeled dataset, the segmentation predictions indicate that for dense and small-scale targets, our FF Swin-Unet outperforms the baseline framework and other models in abdominal CT image segmentation. The segmentation results on unlabeled data have also been validated by radiology experts.

Severity analysis results

To evaluate our proposed automated assessment method for fatty liver grade, we randomly selected 10 original

Table 7 Comparison between the method proposed in this paper and expert manual measurement of the grade of fatty liver cases

Case numbe	Manual measurement	Fatty liver grade	Automatic evaluation	Fatty liver grade
1	0.785	Mild	0.756	Mild
2	0.991	Mild	0.994	Mild
3	1.023	Normal	1.031	Normal
4	0.803	Mild	0.794	Mild
5	0.662	Moderate	0.780	Mild
6	0.873	Mild	0.849	Mild
7	1.04	Normal	1.081	Normal
8	0.477	Severe	0.459	Severe
9	0.996	Mild	0.964	Mild
10	0.939	Mild	0.841	Mild
Total accuracy	90%			

CT scan slices and invited radiology experts to manually measure the liver-spleen ratio. The measurement results were then compared with the automatic analysis results obtained using the FF Swin-Unet model presented in this paper. The manual measurement method by radiologists strictly adhered to clinical protocols, and to minimize errors, the average of three measurements was taken. The liver-spleen ratio was ultimately mapped to the corresponding fatty liver grade within specified ranges. Table 7 presents the results of manual and automated measurements.

According to the analysis in Table 7, the automatic analysis results have a small difference with the average value of manual measurements in the calculation of the specific liver-spleen ratio, especially on the boundary value of 1, but they are basically consistent. However, the difference between the manual measurement result of case number 5 (0.662) and the automatic measurement result (0.780) is large, which may lead to an error in fatty liver grade judgment, possibly due to segmentation boundary error. The last row of the table shows that the overall accuracy of the automatic measurement method is 90%, which can prove the effectiveness of our proposed method to some extent.

Moreover, the Mann-Whitney U test was employed to assess whether there were differences between manual and automated measurements. Based on histogram analysis, it was determined that the shape of the distribution of measurements in both groups was essentially similar. The average for manual measurements was 0.859, with a median of 0.906, while the average for automated measurements was 0.855, with a median of 0.845. The Mann-Whitney U test results indicate that there is no significant difference between manual and automated measurements ($U = 47.000$, $p = 0.853$). The results are summarized in the Table 8.

Table 8 Conclusions of the statistical analysis of the automated severity scoring module (ASSM) results

Methods	Mean	Median level	U	P value
Manual measurement	0.859	0.906	47.000	0.853(> 0.05)
Automatic evaluation	0.855	0.845		

The implications of these results are substantial. The lack of significant difference between manual and automated measurements underscores the high level of accuracy and reliability of the automated system. This suggests that the automated measurements can be confidently used as a surrogate for manual measurements, which is particularly important in clinical settings where time and resources are at a premium. The equivalence in performance also implies that the automated system has the potential to streamline workflow, reduce operator variability, and increase the efficiency of data analysis without compromising on the quality of the results. Furthermore, these results bolster the validity of our model for potential integration into clinical practice, where the consistency and replicability of measurements are critical for patient care and outcome assessment.

Discussion

The development of big data technology and the application of artificial intelligence (AI) algorithms have provided new methods and possibilities for better understanding NAFLD [28]. The advantage of AI over traditional statistical modeling is that it can identify unique patterns and combine multiple factors to create predictive models, risk stratification, and outcomes. It is particularly suitable for chronic diseases because it has heterogeneity, complexity, and overlapping factors [29]. We conducted a retrospective study to validate the independent performance of a new artificial intelligence application for liver and spleen segmentation and the automatic grading of NAFLD severity on abdominal CT images. The data were sourced from The Medical Segmentation Decathlon (MSD) and a high-quality liver and spleen annotation dataset was constructed using the semi-automatization nnU-Net module (SNM). This dataset selection and construction method is crucial for training deep learning models and evaluating their performance. By using a well-annotated dataset, researchers can enhance the accuracy and generalization capability of the model and ensure the reliability of the results.

This paper introduces the Focal Feature Fusion Swin-Unet module (FSM), an NAFLD segmentation detection model aimed at improving the segmentation of the liver and spleen. Quantitative evaluation results indicate that, compared to the original Swin-Unet architecture, our proposed FF Swin-Unet achieves better segmentation performance on evaluation metrics. Specifically, the average DSC is 95.64%, and the average HD95 is 15.94.

Compared to the baseline Swin-Unet method, FF Swin-Unet demonstrates improvements of 0.45% in Average DSC and 1.71 in Average HD95. Compared to existing segmentation methods, the proposed FF Swin-Unet excels in terms of both DSC and HD95. This improvement can be attributed to the design of the main Focal Transformer module and the introduction of the downsampling layer FFM module. These features help the model adapt to cases with significant shape differences and blurred boundaries in the liver and spleen, areas where other models typically fall short.

This paper introduces the automated severity scoring module (ASSM) method to combine with the segmentation model for NAFLD severity grading. By extracting cubes of a certain volume from the liver and spleen and calculating the ratio of CT values, automatic identification and severity assessment of NAFLD can be achieved. Research results demonstrate an accuracy rate of 90% on the test set, highlighting the potential clinical utility of this method. It can alleviate the burden on radiologists and provide patients with more accurate and rapid diagnostic results. Automated NAFLD severity assessment methods can help doctors gain a better understanding of a patient’s condition and guide treatment decisions effectively.

Through a literature review, we found that there were few other similar studies. In the limited relevant studies, other deep learning-based solutions were found to have certain limitations in different aspects compared to the method in this article. Graffy et al. [12] developed a deep learning-based automatic liver fat quantification tool to determine the prevalence of steatosis in a large screening queue using non-contrast CT. By using a three-dimensional convolutional neural network, including a subqueue with subsequent scans, volume-based automatic liver attenuation analysis was analyzed, including conversion to CT fat scores, and compared with manual measurements in a large number of scans. The results showed that this CT-based fully automatic liver fat quantification tool can be used for population-based liver steatosis and NAFLD assessments, and the objective data are in good agreement with manual measurement results. However, the methodological and algorithmic descriptions were fuzzy and could not be practically used by subsequent researchers. Huo et al. [30] proposed an automatic liver attenuation ROI measurement (ALARM) method based on automatic ROI extraction from liver attenuation. It consists of two main stages: (a) liver segmentation based on deep convolutional neural networks (DCNN) and (b) automatic ROI extraction. DCNN combined with morphological operations can achieve “excellent” consistency with manual estimation for detecting fatty liver. The entire pipeline is implemented as a Docker container, but it takes 5 minutes to complete the liver attenuation

evaluation, which is time-consuming. As a radiological tool, in quantifying liver fibrosis, the derivative parameters of image-based texture analysis combined with ML of non-contrast-enhanced T1-weighted magnetic resonance images are similarly accurate as magnetic resonance elastography, but only 82% [11].

The proposed dual-mode deployment framework demonstrates substantial computational efficiency improvements for clinical implementation. Through synergistic optimizations including magnitude-based pruning (18% head removal) and 8-bit quantization (achieving $4.2 \times$ model compression), the system achieves 2.1s/case inference latency on NVIDIA T4 GPUs (95.1% DSC) and 9.8s/case on Xeon CPUs (94.7% DSC), corresponding to a $51\text{--}71 \times$ acceleration compared to manual measurement (1.8–2.5 mins/case). The adaptive memory footprint scaling from 2.1GB (CPU) to 18GB (GPU) supports flexible deployment across varied hospital infrastructures. Clinical validation confirms its reliable performance in real-world NAFLD screening scenarios, including resource-constrained rural clinics without dedicated GPUs. This automated system enhances NAFLD diagnosis by significantly reducing time requirements for segmentation and severity assessment. The rapid generation of accurate severity scores supports timely clinical decision-making and personalized treatment planning, potentially improving patient management. The integration of AI into clinical workflows optimizes radiological resource utilization while maintaining diagnostic quality, particularly valuable for standardizing screening in diverse healthcare settings.

Although this study has yielded promising outcomes, it is not without limitations that warrant further investigation. First, the dataset was derived from the MSD dataset, potentially introducing sample selection bias. Future studies should employ multi-center datasets with expanded sample sizes to validate the model's robustness and generalizability. Second, institutional biases persist in medical imaging data, including residual protocol variations (e.g., 5–10 HU differences due to contrast timing discrepancies) despite preprocessing, reflecting real-world clinical heterogeneity. Third, the single-annotator labeling protocol (where each dataset was annotated by a single radiologist), while mirroring actual clinical workflows, may introduce labeling noise. Furthermore, although high accuracy was achieved in lesion segmentation and severity assessment, there remains room for error reduction. Subsequent work should refine training strategies, expand training datasets, and integrate additional imaging features with clinical data to improve diagnostic precision. Notably, while the automated method alleviates radiologists' workload, physicians' professional judgment and clinical validation remain irreplaceable. Deep learning models should be positioned

as decision-support tools rather than replacements for clinical expertise. To ensure continuous improvement, explicit clinical feedback mechanisms must be established, enabling iterative model optimization based on real-world deployment experiences.

The objective of this research is to engineer a proficient and intuitive automated system for medical professionals. The system interface is designed to facilitate the input of a patient's abdominal CT scans. Subsequently, an automated workflow is initiated, which encompasses programs for the autonomous 3D segmentation of the liver and spleen. This process culminates in the assessment of the non-alcoholic fatty liver disease (NAFLD) severity score by employing multiple regions of interest calculations. This study introduces an integrated Python-API-based platform: NAFLD Severity Scoring System, tailored for abdominal CT imaging data. Upon uploading the patient's abdominal CT scans, the system executes the following automated processes: it employs the FF Swin-Unet algorithm for image enhancement, performs 3D segmentation on the liver and spleen, and computes the NAFLD severity score in accordance with radiological standards. The system not only automatically localizes the liver and spleen within the CT images but also aids clinicians in evaluating the NAFLD severity by integrating the radiological NAFLD severity scoring criteria with the analysis of 3D regional CT value ratios. The system boasts high efficiency, classification accuracy, and maintains a robust level of precision, offering significant utility to the medical professionals.

Conclusions

Utilizing abdominal CT image data, complemented by NAFLD-specific radiological characteristics, we have developed an innovative technique for the segmentation of NAFLD and the automated assessment of disease severity. This method stands on the shoulders of an enhanced Swin-Unet architecture, incorporating a semi-automated dataset construction module (SNM), a Focal Feature Fusion Swin-Unet module (FSM) tailored for precise liver and spleen segmentation, and an Automated Severity Scoring Module (ASSM) for NAFLD grading. Our experimental outcomes demonstrate that this approach outperforms existing state-of-the-art methods in both the segmentation of liver and spleen and the automation of NAFLD severity scoring.

The implications of our study are significant, offering a robust tool for clinicians to not only visualize NAFLD but also to quantitatively assess its severity, which is pivotal for treatment planning and monitoring. Looking ahead, the expansion of our dataset with more labeled images will undoubtedly refine the model's performance. Moreover, the integration of diverse medical imaging modalities has the potential to enhance diagnostic precision

and reliability. To enhance model interpretability, the provision of segmented three-dimensional images will empower practitioners with a clearer understanding of the model's decision-making process. For severity evaluation, while our current approach is based on global liver properties, future research may explore more granular methodologies to refine the assessment's accuracy and dependability.

Acknowledgements

We would thank the support of High-Performance Computing Platform of BML.

Author contributions

All the authors have equally participated as follows: Conceptualization, Zhuowei Zhang and Feng Song; Methodology, Liting Fan and Yi Lei; Validation, Zhuowei Zhang, Feng Song, Liting Fan and Yi Lei; Formal Analysis, Xiangfei Sun; Investigation, Liting Fan, Yi Lei and Xiangfei Sun; Data Curation, Xiangfei Sun; Writing—Original Draft Preparation, Liting Fan and Yi Lei.

Funding

This work was supported by the National Key R&D Program of China (No. 2020YFC2006702 and 2020YFC2005503). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to participate

Since the MSD dataset is derived from publicly available sources and consists of de-identified data, so obtaining explicit informed consent from individual patients was not applicable for this particular study.

Consent for publication

Not applicable.

Clinical trial number

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 March 2025 / Accepted: 24 June 2025

Published online: 10 July 2025

References

- Lazo M, Hernaez R, Eberhardt MS, Bonekamp S, Kamel I, Guallar E, Koteish A, Brancati FL, Clark JM. Prevalence of nonalcoholic fatty liver disease in the United States: The third national health and nutrition examination survey, 1988–1994. *Am J Epidemiol*. 2013;178(1):38–45. <https://doi.org/10.1093/aje/kws448>.
- Adams LA, Anstee QM, Tilg H, Targher G. Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. *Gcut*. 2017;66(6):1138–53. <https://doi.org/10.1136/gutjnl-2017-313884>.
- Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, Swain MG, Congly SE, Kaplan GG, and Shaheen A-A. The prevalence and incidence of nafld worldwide: A systematic review and meta-analysis. *Lancet Gastroenterol. Hepatol*. 2022;7(9):851–61. [https://doi.org/10.1016/S2468-1253\(22\)00165-0](https://doi.org/10.1016/S2468-1253(22)00165-0).
- Lonardo A, Nascimbeni F, Mantovani A, and Targher G. HypertensiHypertensiOn, diabetes, atherosclerosis and nash: Cause or consequence? *J Hepatol*. 2018;68(2):335–52. <https://doi.org/10.1016/j.jhep.2017.09.021>.
- Makri E, Goulas A, Polyzos SA. Epidemiology, pathogenesis, diagnosis and emerging treatment of nonalcoholic fatty liver disease. *Arch Med Res*. 2021;52(1):25–37. <https://doi.org/10.1016/j.arcmed.2020.11.010>.
- Zhou J-H, She Z-G, Li H-L, Cai J-J. Noninvasive evaluation of nonalcoholic fatty liver disease: Current evidence and practice. *World J Gastroenterol*. 2019;25(11):1307–26. <https://doi.org/10.3748/wjg.v25.i11.1307>.
- Li X-H, Liu F, Yao H-Y. New advances of artificial intelligence in the diagnosis of non-alcoholic fatty liver disease. *Zhonghua Gan Zang Bing Za Zhi*. 2022;30(4):4. <https://doi.org/10.3760/cma.j.cn50113-20210806-00380>.
- Perveen S, Shahbaz M, Keshavjee K, Guergachi A. A systematic machine learning based approach for the diagnosis of non-alcoholic fatty liver disease risk and progression. *Sci. Rep*. 2018;8(1). <https://doi.org/10.1038/s41598-018-20166-x>.
- Ji W, Xue M, Zhang Y, Yao H, Wang Y. A machine learning based framework to identify and classify non-alcoholic fatty liver disease in a large-scale population. *Front Public Health*. 2022;10. <https://doi.org/10.3389/fpubh.2022.846118>.
- Che H, Brown LG, Foran DJ, Noshier JL, Hachililoglu I. Liver disease classification from ultrasound using multi-scale cnn. *Int. J. Comput. Assisted Radiol. Surg*. 2021;16(9):1537–48. <https://doi.org/10.1007/s11548-021-02414-0>.
- Huo Y, Terry JG, Wang J, Nair S, Lasko TA, Freedman BI, Carr JJ, Landman BA. Fully automatic liver attenuation estimation combining cnn segmentation and morphological operations. *Med Phys*. 2019;46(8):3508–19. <https://doi.org/10.1002/mp.13675>.
- Graffy PM, Sandfort V, Summers RM, Pickhardt PJ. Automated liver fat quantification at nonenhanced abdominal ct for population-based steatosis assessment. *Radiology*. 2019;293(2):334–42. <https://doi.org/10.1148/radiol.2019190512>.
- Zamanian H, Shalhaf A, Zali MR, Khalaj AR, Dehghan P, Tabesh M, Hatami B, Alizadehsani R, Tan R-S, Acharya UR. Application of artificial intelligence techniques for non-alcoholic fatty liver disease diagnosis: A systematic review (2005–2023). *Comput. Meth. Programs Biomed*. 2024;244:107932.
- Li Y-F, Wang X-T, Zhang J, Zhang S-S, Jiao J. Applications of artificial intelligence (ai) in researches on non-alcoholic fatty liver disease (nafld): A systematic review. *Rev Endocr Metab Disord*. 2022;23(3):387–400. <https://doi.org/10.1007/s11154-021-09681-x>.
- Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, Litjens G, Menze B, Ronneberger O, Summers RM, van Ginneken B, Bilello M, Bilic P, Christ PF, Do RKG, Gollub MJ, Heckers SH, Huisman H, Jarnagin WR, McHugo MK, Napel S, Pernicka JS, Rhode K, Tobon-Gomez C, Vorontsov E, Meakin JA, Ourselin S, Wiesenfarth M, Arbeláez P, Bae B, Chen S, Daza L, Feng J, He B, Isensee F, Ji Y, Jia F, Kim I, Maier-Hein K, Merhof D, et al. The medical segmentation decathlon. *Nat Commun*. 2022;13(1). <https://doi.org/10.1038/s41467-022-30695-9>.
- Ya QCY. The segmented datasets of liver CT and spleen CT. 2023. <https://aistudio.baidu.com/aistudio/datasetdetail/194126>, Feb 26.
- Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M. Swin-Unet: Unet-like pure transformer for medical image segmentation. *Lect Notes Comput Sci*. 2023;13803 LNCS:205–18. <https://doi.org/10.1007/978-3-031-25066-89>.
- Yang J, Li C, Zhang P, Dai X, Xiao B, Yuan L, Gao J. Focal attention for long-range interactions in vision transformers. 2021;36:30008–22.
- Yu C, Gao C, Wang J, Yu G, Shen C, Sang N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *Int J Comput Vision*. 2021;129(11):3051–68. <https://doi.org/10.1007/s11263-021-01515-2>.
- Liu Y, Chu L, Chen G, Wu Z, Chen Z, Lai B, Hao Y. Paddleseg: A high-efficient development toolkit for image segmentation. 2021.
- Jeong WK. Hepatic and splenic volumetry could be used as an imaging parameter to evaluate fibrosis grades of the diffuse liver disease including nonalcoholic fatty liver disease. *Gut Liver*. 2017;11(5):577–78. <https://doi.org/10.5009/gnl17333>.
- Zhang DF, Lu W, Mi YQ, Li P. Diagnostic value of liver and spleen ct ratio on the degree of steatosis in nonalcoholic fatty liver disease. *Chin Hepatol*. 2012;17(4):252–53. <https://doi.org/10.3969/j.jissn.1008-1704.2012.04.009>.
- Chen J, Mei J, Li X, Lu Y, Yu Q, Wei Q, Luo X, Xie Y, Adeli E, Wang Y, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Med Image Anal*. 2024;97:103280. <https://doi.org/10.1016/j.media.2024.103280>.

24. Zhou H-Y, Guo J, Zhang Y, Han X, Yu L, Wang L, Yu Y. Nnformer: Volumetric medical image segmentation via a 3d transformer. *IEEE Trans Image Process*. 2023;32:4036–45. <https://doi.org/10.1109/TIP.2023.3293771>.
25. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. *ArXiv Abs/2201.01266*. 2022.
26. Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016;565–71. <https://doi.org/10.1109/3DV.2016.79>.
27. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. Nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203–11. <https://doi.org/10.1038/s41592-020-01008-z>.
28. Anand V, Gupta S, Altameem A, Nayak SR, Poonia RC, Saudagar AKJ. An enhanced transfer learning based classification for diagnosis of skin cancer. *Diagn*. 2022;12(7). <https://doi.org/10.3390/diagnostics12071628>.
29. Dinani AM, Kowdley KV, Nouredin M. Application of artificial intelligence for diagnosis and risk stratification in nafld and nash: The state of the art. *Hepatology*. 2021;74(4):2233–40. <https://doi.org/10.1002/hep.31869>.
30. Qu H-Q, Li Q, Grove ML, Lu Y, Pan J-J, Rentfro AR, Bickel PE, Fallon MB, Hanis CL, Boerwinkle E, McCormick JB, Fisher-Hoch SP. Population-based risk factors for elevated alanine aminotransferase in a south texas Mexican-American population. *Arch Med Res*. 2012;43(6):482–88. <https://doi.org/10.1016/j.jarcmed.2012.08.005>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.