A Unified Registration-Guided 3DINO Framework for Semi-Supervised Liver Segmentation and Patch-Based Liver Fibrosis Staging under Missing Modalities

Boya Wang[1⋆], Ruizhe Li[1,2], Chao Chen[1] and Xin Chen[1]

[1] School of Computer Science, University of Nottingham, UK
[2] Nottingham Biomedical Research Centre (BRC), School of Medicine, University of Nottingham, UK

Abstract. Precise liver segmentation and fibrosis staging are vital for clinical diagnosis and treatment planning in liver fibrosis, representing a major challenge in clinical practice. However, Multi-parametric MRI often suffers from missing imaging modalities. To address this, we propose a registration-guided semi-supervised liver segmentation and 3D patch-based classification framework for the CARE 2026 Challenge. The segmentation component is designed for GED4, T1, T2, and DWI, while the classification component supports selected modalities and modality combinations under incomplete availability. Extensive evaluation on the CARE Liver 2026 test set demonstrates that our unified framework achieves highly robust performance in both liver segmentation and fibrosis staging. The findings indicate that the proposed framework demonstrates proficiency in effectively managing missing modalities, limited labels, and domain shifts. The code is available on this GitHub link: https://github.com/mileywang3061/Care-Liver2026

Keywords: Liver fibrosis · Multi-parametric MRI · 3DINO · Patch-based classification.

1 Introduction

Liver fibrosis is a critical intermediate stage in the progression of chronic liver diseases, such as viral hepatitis, fatty liver, and alcoholic liver disease, toward cirrhosis and liver cancer [1, 13]. As a direct quantitative indicator of the progression of chronic liver diseases, it guides clinical decision-making and determines the prognosis. The CARE-Liver challenge [2] aims to develop novel and effective AI-based methods for liver segmentation and fibrosis staging using multi-phase, multi-vendor MRI data. In this challenge, fibrosis severity is represented by four ordered categories, S1–S4, with S4 corresponding to cirrhosis.

This study addresses both liver segmentation and liver fibrosis staging. The classification model is constructed based on the output from the segmentation

⋆ Corresponding author: boya.wang@nottingham.ac.uk

process. For the segmentation task, a registration-guided semi-supervised approach has been adopted, as only a small fraction of one phase carries manual annotations, while all other phases are used without labels. Semi-supervised methods, such as the Mean Teacher [11] and pseudo tag [3, 7], are able to utilize unlabeled data. However, difficulties have been encountered in maintaining consistency across modality space and structure when employing them [7]. In contrast, the joint registration–segmentation frameworks [5, 8, 20] have demonstrated that the enforcement of geometric and label consistency through explicit spatial alignment is a promising approach.

For the liver fibrosis staging task, early studies largely relied on radiomics, which is widely utilized to extract high-dimensional quantitative features, such as texture, intensity, and shape [12, 17]. Although these approaches demonstrated the potential of quantitative MRI analysis, their dependence on accurate lesion or liver segmentation and manually designed features limited their scalability, reproducibility, and robustness across imaging protocols and clinical centers. With the accelerated development of deep learning, convolutional neural networks (CNNs) enabled end-to-end representation learning directly from medical images and subsequently became a dominant approach for automated liver fibrosis assessment [21]. 2D patch-based learning strategies for multi-parametric MRI, such as the method proposed by Wang et al. [14, 16], have achieved promising performance. However, 2D blocking may compromise the spatial information relationship between slices and generate discontinuous and inconsistent cross-volume representation, thereby compromising the integrity of the data. Furthermore, the provision of contextual information is limited, and volume patterns related to liver morphology and fibrosis distribution may be ignored.

Consequently, this study investigates 3D block-based methods as a primary research direction. 3DINO [19], which is fundamentally built on the robust DINOv2 paradigm by strategically fine-tuning, is a generalizable 3D self-supervised learning framework for medical imaging. 3DINO learns highly adaptive 3D representations without the need for manual annotation through further self-supervised pretraining on different medical datasets. Therefore, it can serve as an exceptional feature extractor for 3D frameworks, ensuring stable diagnostic performance even when MRI modalities are missing.

This study proposes a unified framework to jointly achieve accurate liver segmentation and fibrosis staging from multi-phase and multi-center liver MRI data with modality missing. Specifically, in the segmentation part, we combine registration-guided semi-supervised learning, a pre-trained 3DINO-ViT architecture, and a cascaded nnU-Net refinement module to effectively overcome the bottleneck of limited annotations through cross-modal label propagation. In the classification part, the framework adopts a three-dimensional image patch strategy to extract fine-grained features using the 3DINO-ViT architecture, and utilizes the stage 1 and 4 image patches as proxies for health and lesion features, respectively. The spatial proportion of high-risk image patches (S4) is used for patient-level stage inference.

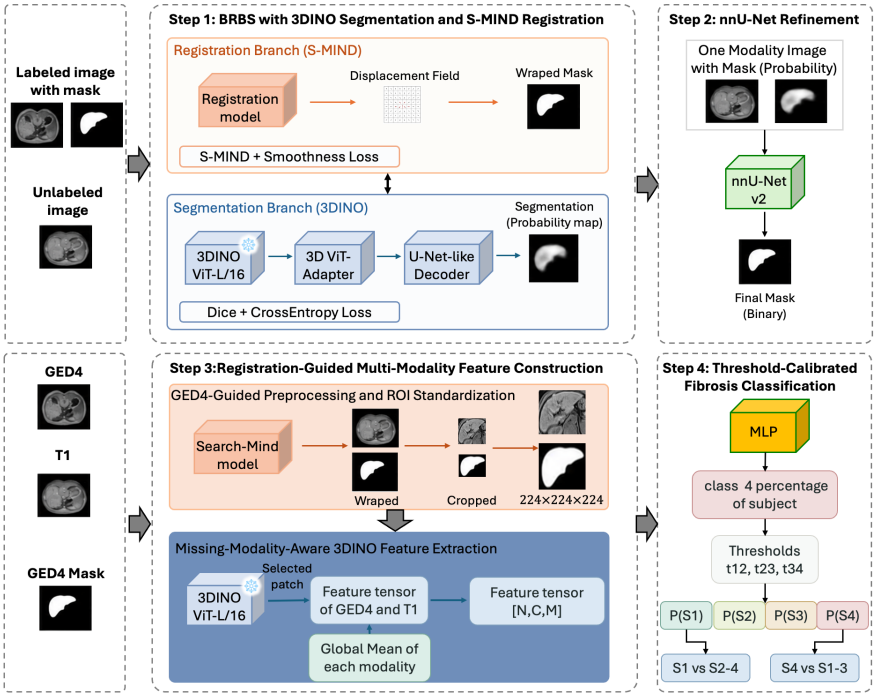

Fig. 1: Overview of the proposed framework. In the segmentation branch, labeled and unlabeled images are processed by BRBS with S-MIND-based registration and a 3DINO-ViT segmentation branch to generate initial probability maps, which are independently refined by nnU-Net. In the classification branch, the MRI modalities and mask are aligned to GED4, cropped using the liver mask, resized to 224^3, and encoded by a pretrained 3DINO-ViT-L/16. The resulting missing-modality-aware patch features are classified by an MLP for liver fibrosis staging.

2 Methods

This section presents the proposed framework for semi-supervised liver segmentation and subject-level liver fibrosis staging under missing modalities. As illustrated in Fig. 1, the framework consists of a registration-guided segmentation pipeline combining BRBS, S-MIND, a pretrained 3DINO-ViT segmentation branch, and nnU-Net refinement, together with a missing-modality-aware 3D patch-based classification pipeline.

2.1 Registration-guided Semi-supervised Liver Segmentation

The segmentation pipeline consists of a BRBS-based semi-supervised step with S-MIND registration and 3DINO-ViT segmentation, followed by nnU-Net refinement.

BRBS with S-MIND and 3DINO: The first step is built upon the BRBS framework [5]. Its registration branch employs a 3D U-Net that takes the concatenated moving and fixed images as input and predicts a dense displacement field. A differentiable spatial transformer then warps the labeled liver mask into the space of an unlabeled target image, providing a registration-derived pseudo-label for semi-supervised segmentation. The Weighted Consistency Constraint and Space-Style Sampling Program of BRBS are retained to reduce the influence of unreliable propagated labels and to generate additional training samples with intermediate anatomical deformations and MRI appearances.

To improve the robustness of the BRBS **registration branch** across MRI modalities, the original NCC similarity term is replaced by the S-MIND loss proposed in Search-MIND [15]. The BRBS registration architecture remains unchanged, while S-MIND is used as the image-similarity objective during training. By comparing modality-independent neighbourhood descriptors within a local search window, S-MIND is more robust to nonlinear intensity relationships and residual spatial displacement across MRI sequences. A deformation smoothness loss is additionally applied to encourage anatomically plausible displacement fields.

The original BRBS **segmentation branch** is replaced by the downstream segmentation architecture introduced with 3DINO [19]. It consists of a pretrained 3DINO-ViT-L/16 encoder, a 3D ViT-Adapter, and a U-Net-like decoder. The 24 Transformer layers are divided into four blocks, and their representations are converted by the ViT-Adapter into a four-level feature pyramid. These features are progressively upsampled by decoder layers with 256, 128, 64, and 32 channels, together with encoder–decoder skip connections. The pretrained 3DINO-ViT encoder is kept frozen throughout training, while the ViT-Adapter and decoder are optimized using manual annotations and registration-derived pseudo-labels with Dice and cross-entropy losses. Freezing the encoder reduces the number of trainable parameters and limits overfitting to the small labeled dataset and potentially noisy propagated masks.

nnU-Net Refinement: Finally, the initial segmentation probability map generated for each MRI sequence is treated as an independent single-channel sample and processed by a shared second-step nnU-Net v2 using the 3D full-resolution configuration [6]. Predictions from different modalities are therefore refined independently rather than being stacked as fixed input channels. This design allows the same refinement model to process any available MRI sequence without requiring a complete set for each subject. The nnU-Net is trained using the corresponding manual or registration-derived liver masks to correct inaccurate boundaries, disconnected components, and residual errors from the first-step segmentation. Each refined prediction is converted into a binary liver mask and restored to the original image geometry. The aligned multi-parametric MRIs and refined liver masks are subsequently used for liver ROI extraction and patch-based fibrosis staging.

2.2 3D Patch-based Liver Fibrosis Staging

The classification framework consists of registration-guided multi-modality feature construction and threshold-calibrated fibrosis classification in two steps for subject-level liver fibrosis staging from heterogeneous multi-center, multi-phase MRI data with incomplete modality availability. The overview of the workflow is shown in Fig 1 (step 3 and 4).

GED4-Guided Preprocessing and ROI Standardization: For each subject, the GED4 modality was used as the anatomical reference. Initially, GED4 was resampled to an isotropic voxel spacing of $1 \times 1 \times 1$ mm. All other available MRI sequences were then resampled onto the GED4 reference grid to ensure consistent image size, voxel spacing, origin, and orientation. After grid standardization, each non-reference modality was rigidly registered to GED4 by using the Search-MIND [15] registration method to correct for residual inter-sequence misalignment caused by patient motion and respiratory variation. After spatial standardization and inter-modality registration, the liver ROI was extracted by using the corresponding liver segmentation mask. Since all modalities have been registered to GED4, the GED4 mask after grid standardization would be regarded as the mask of the liver for all the modalities. To ensure cover the whole liver, the bounding box was expanded by 5 pixels along every spatial axis and then applied consistently to the registered MRI modalities and the liver mask to obtain the liver ROI. The cropped liver ROI of each available modality was resized to a fixed volume size of $224 \times 224 \times 224$ and independently fed into a pretrained 3DINO-ViT encoder.

Missing-Modality-Aware 3DINO Feature Extraction: According to the structure of the 3DINO-ViT encoder, the normalized patch tokens with a patch size of $16 \times 16 \times 16$, a volume of $14 \times 14 \times 14$ grid, were obtained from the encoder output. In parallel, the corresponding liver mask was resized to the same spatial size and used to calculate the liver occupancy of each patch. Patch selection was restricted to tokens whose corresponding mask occupancy was greater than 70%, ensuring that only liver-dominant regions contributed to the downstream representation. The retained patch tokens from all available modalities were then stacked to construct a subject-level multi-modality feature pool of size $N \times C \times M$, where N is the number of selected liver patches, C is the 3DINO feature dimension, and M is the number of modalities. For this task, 7 channels are considered: T1, T2, DWI, GED1, GED2, GED3, and GED4. Missing modalities are represented by NaN-value features and subsequently replaced with modality-specific global mean features.

During training, in order to address the issue of incomplete modality protocols, missing-modalities handling was incorporated. Specifically, the dropout of online modality was applied to simulate incomplete availability of modality. For each training sample, up to one available modality was randomly removed from the T1/T2/DWI group and up to one from the GED1/GED2/GED3 group, while

GED4 was always retained. The dropped modalities were also replaced with their corresponding global mean features. During validation and testing, no random dropout was applied, and only truly missing modalities were mean-filled.

Threshold-Calibrated Fibrosis Classification: The implementation of a patch-level fibrosis classifier utilizing a multilayer perceptron on concatenated multi-modality 3DINO features was conducted. The model was formulated as an extreme-stage binary classification task, using S1 and S4 cases as good and bad part references, respectively. The optimization process incorporated balanced patch sampling and cross-entropy loss. In intermediate-stage cases, validation-based calibration was utilized as the primary method of calibration, superseding the conventional approach of direct binary supervision.

The subject-level fibrosis staging was obtained by averaging the patch-level Stage-4-like probabilities into a continuous class-4 percentage. Three calibrated thresholds, which were gotten from the validation, were used to map this score to four fibrosis stages: t_{12} for S1 versus S2, t_{23} for S2 versus S3, and t_{34} for S3 versus S4. Here is the formula of fibrosis probability mapping:

1. The sigmoid steepness parameters were defined as:

$$k_{12} = \frac{4}{t_{23}}, \qquad k_{23} = \frac{4}{t_{34} - t_{12}}, \qquad k_{34} = \frac{4}{1 - t_{23}} \tag{1}$$

   The selection of a factor of 4 in sigmoid steepness parameters is to guarantee that the majority of the sigmoid transition occurs within the corresponding threshold interval. When ($k = 4/d$), the sigmoid output changes from approximately 0.12 to 0.88 over an interval of d.
2. The cumulative probabilities were computed using sigmoid functions:

$$\begin{aligned} &f_{12}(x) = \sigma(k_{12}(x - t_{12})), \quad f_{23}(x) = \sigma(k_{23}(x - t_{23})), \\ &f_{34}(x) = \sigma(k_{34}(x - t_{34})), \quad \text{where: } \sigma(z) = \frac{1}{1 + e^{-z}} \end{aligned} \tag{2}$$

3. The four-stage probabilities were derived as:

$$\begin{aligned} &P(S1) = 1 - f_{12}(x), \qquad P(S2) = f_{12}(x) - f_{23}(x), \\ &P(S3) = f_{23}(x) - f_{34}(x), \quad P(S4) = f_{34}(x) \end{aligned} \tag{3}$$

The final predictions of S1 vs S2-S4 and S4 vs S1-S3 were generated using $P(S1)$ and $P(S4)$.

3 Experiments and Results

3.1 Dataset

The **CARE-Liver** dataset [10, 9, 4, 18] comprises multi-parametric MRI scans from 800 patients, collected across multiple clinical centers by employing one of

Table 1: Summary of dataset splits in CARE-Liver. Values are shown as total cases (labeled liver mask cases).

Dataset	A	B1	B2	OOD (C)	Total
Training	157 (10)	230 (10)	73 (10)	–	460 (30)
Validation	20 (20)	20 (20)	20 (20)	–	60 (60)
Test	50 (50)	50 (50)	50 (50)	130 (130)	280 (280)

three scanner models: Philips Ingenia (3.0T), Siemens Skyra (3.0T), or Siemens Aera (1.5T). The record of each patient includes a subset of sequences: T1-weighted (T1), T2-weighted (T2), diffusion-weighted imaging (DWI), and dynamic phases enhanced with Gd-EOB-DTPA (GED1-GED4), while the GED4 phase is invariably available for all subjects. The distribution of this dataset under training, validation, and test in three centers and four vendors A, B (B1, B2), and C is shown in Table 1.

The training set contains 460 cases distributed in different fibrosis stages: 99 in Stage 1, 75 in Stage 2, 48 in Stage 3, and 238 in Stage 4. Within this set, each vendor of A, B1, and B2 provides 10 manually annotated liver masks of the GED4 phase. The validation set includes 60 cases (20 per vendor). The test set comprises 280 cases: 50 from each in-distribution group, while 130 out-of-distribution (OOD) cases from vendor C and center C were specifically utilized for the evaluation of the model's generalization capabilities.

3.2 Parameter Setting and Experiment Setting

All trainable components, except for the pretrained 3DINO-ViT-L/16 encoder, were trained from scratch using a learning rate of 1×10^{-4}. The segmentation network was trained for 500 epochs with all available data, and each epoch contains 200 iterations. The patch-level classifier was implemented with an MLP with an input dimension of 1024×7, followed by hidden layers of 2048, 512, and 128 units and a two-class output layer. Dropout was set to 0.3. The classifier was trained for 30 epochs using cross-entropy loss and the Adam optimizer with a weight decay of 1×10^{-4}. A StepLR scheduler with a step size of 5 and a decay factor of 0.7 and the batch size was 256.

The classification experiment protocol consisted of two consecutive cross-validation procedures. Firstly, a repeated nested validation procedure with 10 iterations and 4 folds was conducted for the fixed threshold. In each iteration, the training dataset was stratified proportionally, with 90% of the samples designated for development and 10% for retention testing. Within the development set, 4-fold cross-validation was performed using the S1 and S4 subjects. In each fold, three subsets of the S1/S4 samples were used to train the binary classifier, while the remaining S1/S4 subset, together with all S2 and S3 cases in the development set, formed the validation set. The model training is based exclusively on the S1 and S4 extreme cases. At the same time, the validation set encompasses

the entire stage data from S1 to S4, thereby facilitating the precise estimation of the thresholds t_{12}, t_{23}, and t_{34}. Subsequently, 10% of the test set is retained in each iteration for the purpose of evaluating the calibrated ensemble model. Finally, the thresholds calculated from multiple iterations are aggregated to define the final fixed threshold set. Then, a final 4-fold cross-validation was conducted with all training data with the fixed threshold set. The best checkpoint from each fold was retained and combined into a final ensemble for inference.

3.3 Segmentation Results

For the liver segmentation subtask (LiSeg), we first evaluated the proposed framework on the validation set using GED4, T1, T2, and DWI images. Three configurations were compared to assess the contributions of the multimodal registration loss, the pretrained 3DINO-ViT segmentation branch, and the cascaded nnU-Net refinement stage.

We compared three configurations: (1) **BRBS-NCC**, the original BRBS framework using Normalized Cross-Correlation as the registration similarity loss; (2) **BRBS-S-MIND**, in which NCC is replaced by the S-MIND loss for multimodal deformable registration; and (3) **Proposed**, the complete two-stage framework combining S-MIND-based BRBS, a frozen pretrained 3DINO-ViT segmentation encoder, and nnU-Net refinement. All configurations were trained and evaluated using the same data split.

As shown in Table 2, replacing NCC with S-MIND consistently improved both Dice and HD across all four MRI sequences. The Dice improvements were most evident on T2 and DWI, increasing from 0.7750 to 0.7947 and from 0.8094 to 0.8189, respectively. S-MIND also substantially reduced the DWI HD from 53.34 to 39.31. These results indicate that the local descriptor search provides more reliable registration and pseudo-label propagation under cross-modality intensity differences.

Table 2: Liver segmentation results on the CARE Liver 2026 validation set.

Method	**GED4**		**T1**		**T2**		**DWI**	
	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓
BRBS-NCC	0.9528	26.73	0.9316	65.70	0.7750	63.27	0.8094	53.34
BRBS-S-MIND	0.9560	**21.70**	0.9328	60.48	0.7947	**57.55**	0.8189	**39.31**
Proposed	**0.9689**	26.30	**0.9415**	**58.62**	**0.8447**	65.55	**0.8463**	59.08

The proposed framework achieved the highest Dice score for every modality. Compared with BRBS-S-MIND, the largest Dice improvement was observed on T2, with an increase of 0.0500, followed by DWI with an increase of 0.0274. The

mean Dice across the four modalities increased from 0.8756 for BRBS-S-MIND to 0.9004 for the proposed framework.

However, the improvements in Dice were not consistently accompanied by lower HD. The proposed method achieved the lowest HD on T1, whereas BRBS-S-MIND retained lower HD values on GED4, T2, and DWI. This suggests that the 3DINO-based segmentation and nnU-Net refinement improved overall volumetric overlap but occasionally introduced isolated boundary errors or distant false-positive regions, to which the Hausdorff distance is particularly sensitive.

Table 3: Official CARE Liver 2026 test-set results for liver segmentation. In-distribution and out-of-distribution cases are reported separately.

Modality	**ID**		**OOD**	
	Dice↑	HD↓	Dice↑	HD↓
GED4	0.9665	33.60	0.9749	22.80
T1	0.9490	45.00	0.9521	28.24
T2	0.8558	64.25	0.8495	44.28
DWI	0.8564	52.08	0.9128	25.43
Average	0.9069	48.73	0.9223	30.19

On the official test set, the proposed framework achieved consistently strong segmentation performance on both ID and OOD cases. The average Dice scores were 0.9069 and 0.9223 for ID and OOD data, respectively, with average HD values of 48.73 and 30.19. No clear performance degradation was observed on the OOD cases, and DWI showed a particularly noticeable improvement in Dice from 0.8564 to 0.9128. Overall, these results indicate that the framework maintains stable segmentation performance across different test distributions.

3.4 Classification Results

Based on the segmentation results, we further evaluated the classification performance of liver fibrosis staging. For comparison, DINO-Med [14] was specifically selected on the basis that the DINO architecture is also employed for the purpose of feature extraction, thus providing a fair and direct baseline with which to evaluate the performance improvements achieved by expanding to our pipeline. Additionally, three combinations of modalities were evaluated based on our framework: 7-Modality (T1, T2, DWI, GED1, GED2, GED3, and GED4), 4-Modality (T1, T2, DWI, GED4), and 1-Modality (GED4).

Table 4 shows the performance of our proposed framework across different modality availability scenarios and the 2D DINO-Med framework in the 7-Modality. It is noteworthy that the models demonstrated varying sensitivities to modality richness depending on the specific task. Concerning the detection

Table 4: Liver fibrosis staging results on the CARE Liver 2026 validation set.

Modality	**Task 1 (S1-S3 vs S4)**		**Task 2 (S1 vs S2-S4)**	
	ACC↑	AUC↑	ACC↑	AUC↑
3D-7-Modality	**0.7333**	0.7715	0.7000	0.7363
3D-4-Modality	0.6833	**0.7823**	0.7000	0.7074
3D-1-Modality	0.7000	0.7404	0.7000	0.7030
2D-7-Modality	**0.7333**	0.7679	**0.7333**	**0.7637**

of late-stage fibrosis (Task 1), the 4-Modality configuration achieved the highest Area Under Curve (AUC) value of 0.7823. This outcome indicates that the core set of modalities in question offers sufficient features, whilst other modalities may introduce minor redundancies concerning feature contribution. However, for the early fibrosis (Task 2: S1 vs. S2–S4) detection, the incorporation of all seven modalities proved to be essential. While the accuracy (ACC) remained consistent at 0.7000, the seven-modality configuration enhanced the AUC to 0.7363, signifying that integrating multi-parametric MRI substantially improves the model's capacity to identify and predict early borderline cases with enhanced confidence. For the 2D DINO-Med framework, both task 1 and task 2 achieved the best ACC at 0.7333. Especially, for AUC in task 2, it got the best performance at 0.7637 of all evaluations.

Table 5: Official CARE Liver 2026 test-set results for liver fibrosis staging on in-distribution (ID) and out-of-distribution (OOD) cases. Task 1: S1-S3 vs S4; Task 2: S1 vs S2-S4. 7-Modality inputs use all available modalities.

Modality	**ID**				**OOD**			
	Task 1		**Task 2**		**Task 1**		**Task 2**	
	ACC↑	AUC↑	ACC↑	AUC↑	ACC↑	AUC↑	ACC↑	AUC↑
3D-7-Modality	0.8133	**0.8806**	**0.8467**	**0.9064**	0.5769	**0.8044**	**0.8462**	**0.7170**
2D-7-Modality	0.8133	0.8682	0.8267	0.8571	**0.6385**	0.7062	0.8154	0.5945

The proposed 3D framework and the 2D DINO-Med baseline were subsequently evaluated on the official CARE Liver 2026 test set under all modality configurations. As illustrated in Table 5, the 3D framework demonstrates a significant advantage over the 2D baseline. On the ID dataset, the 3D framework demonstrated exceptional prediction confidence, achieving 0.8133 in ACC and 0.8806 in AUC in Task 1, in addition to an ACC of 0.8467 and a notable AUC of 0.9064 for Task 2, consistently outperforming the 2D baseline. For the OOD

dataset, although the 3D model yielded a lower ACC(0.5769), it still maintained a higher AUC than the 2D model. Furthermore, in Task 2, the 3D model sustained a higher performance in both ACC and AUC. These tests indicate that, by preserving the continuous spatial topology along the depth axis, the 3DINO-based volume representation extracts inherently robust features, effectively overcoming the limitations of discontinuous cross-volume processing inherent in two-dimensional, slice-by-slice processing.

4 Conclusion

This study introduces a unified registration-guided framework for semi-supervised liver segmentation and 3D patch-based liver fibrosis staging under missing modalities. The segmentation pipeline combines S-MIND-based multimodal registration, a pretrained 3DINO-ViT segmentation model, and nnU-Net refinement. Validation results showed that S-MIND improved multimodal segmentation, while the complete pipeline further increased Dice performance across GED4, T1, T2, and DWI. However, the Hausdorff Distance did not consistently improve, suggesting that additional post-processing, such as connected-component removal and boundary smoothing, may be beneficial. The classification pipeline generates a 3D patch-based classifier based on a pretrained 3DINO-ViT encoder, with modality missing. The test results illustrate that our 3D framework outperforms the 2D DINO-Med method across all ID data tasks, demonstrating greater predictive confidence and feature extraction capabilities. More notably, when faced with out-of-distribution (OOD) data, the 3D model successfully withstands the severe challenge of domain shift by preserving the continuous three-dimensional spatial topology, thereby completely avoiding the catastrophic performance collapse encountered by 2D slice-by-slice processing methods.

References

1. Berzigotti, A., Tsochatzis, E., Boursier, J., Castera, L., Cazzagon, N., Friedrich-Rust, M., Petta, S., Thiele, M.: Easl clinical practice guidelines on non-invasive tests for evaluation of liver disease severity and prognosis–2021 update. Journal of hepatology **75**(3), 659–689 (2021)
2. CARE 2026 Organizing Committee: CARE 2026 Track Liver. `https://zmic.org.cn/care_2026/track_liver/`, accessed: 2026-07-25
3. Deng, G., Sun, H., Xie, W.: Correlation-based switching mean teacher for semi-supervised medical image segmentation. Neurocomputing **633**, 129818 (2025)
4. Gao, Z., Liu, Y., Wu, F., Shi, N., Shi, Y., Zhuang, X.: A reliable and interpretable framework of multi-view learning for liver fibrosis staging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 178–188 (2023)
5. He, Y., Ge, R., Qi, X., Chen, Y., Wu, J., Coatrieux, J.L., Yang, G., Li, S.: Learning better registration to learn better few-shot medical image segmentation: Authenticity, diversity, and robustness. IEEE Transactions on Neural Networks and Learning Systems **35**(2), 2588–2601 (2022)

6. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
7. Jia, D.: Semi-supervised multi-organ segmentation with cross supervision using siamese network. In: MICCAI Challenge on Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation, pp. 293–306. Springer (2022)
8. Li, R., Figueredo, G., Auer, D., Dineen, R., Morgan, P., Chen, X.: A unified framework for semi-supervised image segmentation and registration. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2025)
9. Liu, Y., Gao, Z., Shi, N., Wu, F., Shi, Y., Chen, Q., Zhuang, X.: Merit: Multi-view evidential learning for reliable and interpretable liver fibrosis staging. Medical Image Analysis **102**, 103507 (2025)
10. Liu, Y., Shi, N., Zhang, Z., Zhang, H., Wang, B., Yu, D., Wang, N., Jin, Y., Zhou, Y., Yuan, K., Wang, S., Yang, L., Qiao, X., Liu, W., He, X., Hong, X., Zheng, G., Chen, X., Yang, G.Z., Zhang, L., Li, L., Shi, Y., Zhuang, X.: How far has ai come in liver fibrosis staging? a large-scale real-world dataset and benchmark (2026), `https://arxiv.org/abs/2605.25595`
11. Lou, Q., Lin, T., Qian, Y., Lu, F.: Semi-supervised liver segmentation based on local regions self-supervision. Medical Physics **51**(5), 3455–3463 (2024)
12. Park, H.J., Lee, S.S., Park, B., Yun, J., Sung, Y.S., Shim, W.H., Shin, Y.M., Kim, S.Y., Lee, S.J., Lee, M.G.: Radiomics analysis of gadoxetic acid–enhanced mri for staging liver fibrosis. Radiology **290**(2), 380–387 (2019)
13. Schuppan, D., Afdhal, N.H.: Liver cirrhosis. The Lancet **371**(9615), 838–851 (2008)
14. Wang, B., Li, R., Chen, C., Chen, X.: Dino-med: A unified patch-based adaptation framework for multi-modal medical image analysis applied to liver fibrosis staging. In: International Conference on AI in Healthcare. pp. 122–135. Springer (2026)
15. Wang, B., Li, R., Chen, C., Chen, X.: Search-mind: Training-free multi-modal medical image registration. arXiv preprint arXiv:2604.09743 (2026)
16. Wang, B., Li, R., Chen, C., Chen, X.: Semi-supervised liver segmentation and patch-based fibrosis staging with registration-aided multi-parametric mri. arXiv preprint arXiv:2602.09686 (2026)
17. Wei, J., Jiang, H., Gu, D., Niu, M., Fu, F., Han, Y., Song, B., Tian, J.: Radiomics in liver diseases: Current progress and future opportunities. Liver International **40**(9), 2050–2063 (2020)
18. Wu, F., Zhuang, X.: Minimizing estimated risks on unlabeled data: A new formulation for semi-supervised medical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(5), 6021–6036 (2023)
19. Xu, T., Hosseini, S., Anderson, C., Rinaldi, A., Krishnan, R.G., Martel, A.L., Goubran, M.: A generalizable 3d framework and model for self-supervised learning in medical imaging. npj Digital Medicine **8**(1), 639 (2025)
20. Xu, Z., Niethammer, M.: Deepatlas: Joint semi-supervised learning of image registration and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 420–429. Springer (2019)
21. Yasaka, K., Akai, H., Kunimatsu, A., Abe, O., Kiryu, S.: Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid–enhanced hepatobiliary phase mr images. Radiology **287**(1), 146–155 (2018)