SLICE-MIL: Counterfactual Semantic Anchoring of Disentangled Evidence for Child-Pugh Grading

Yihan Deng[1⋆], Hongwei Huang[2⋆], Yunsong Peng[3], Lei He[1], Wenju Cui[1], Yilin Leng[1], Shuai Wang[4], Qi Cheng[2✉], and Jian Zheng[1✉]

[1] School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China
[2] Department of hepato-pancreato-biliary, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[3] Guizhou Province International Science and Technology Cooperation Base for Precision Imaging Diagnosis and Treatment, Department of Radiology, Guizhou Provincial People's Hospital, Guizhou, China
[4] Department of Hepatobiliary Surgery, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou

Abstract. Child-Pugh (CP) grading, a widely used clinical standard for therapeutic decision-making in hepatocellular carcinoma, integrates biochemical and clinical markers to assess hepatic reserve. Estimating CP grade from routinely acquired contrast-enhanced CT (CECT) may provide opportunistic, complementary imaging evidence and facilitate investigation of associations between structural imaging phenotypes and clinically assessed hepatic reserve. However, inferring CP-related functional status from structural imaging requires overcoming a fundamental structure–function semantic gap. Because CP grade is determined by multiple biochemical and clinical factors, its imaging manifestations are heterogeneous and deeply entangled. Conventional Multiple Instance Learning (MIL) architectures do not explicitly decouple these complex representations, making them susceptible to associative shortcut learning rather than capturing stage-relevant pathological patterns. To explicitly model this structure-function mapping, we propose SLICE-MIL, a spatial-logical framework for ordinal evidence disentanglement. Structurally, an anatomical mask-guided mechanism explicitly decouples the entangled intra- and extra-hepatic features. Functionally, a hierarchical evidence disentanglement module employs representation-level counterfactual ablations to anchor stage-specific semantics, encouraging the decoupled representations to align with the progression logic of CP deterioration. Multicenter validation demonstrates that explicitly bridging this semantic gap mitigates reliance on spurious correlations, substantially improving cross-domain robustness and interpretability. The code is publicly available at https://github.com/cutecattutu/SLICE-MIL.

Keywords: Child-Pugh Grading · Multiple Instance Learning · Structure-Function Mapping · Semantic Anchoring · Feature Disentanglement

⋆: Equal contribution; ✉: chengqi@hust.edu.cn,zhengj@sibet.ac.cn

1 Introduction

Accurate assessment of hepatic reserve guides therapeutic decisions in hepatocellular carcinoma management [1]. As the clinical standard for risk stratification [2], the Child-Pugh (CP) grade is a multi-source systemic metric integrating biochemical markers (albumin, bilirubin, and prothrombin time) and clinical symptoms (ascites and encephalopathy). Consequently, the visual manifestations of CP decompensation on contrast-enhanced CT (CECT) are heterogeneous and entangled, encompassing localized intra-hepatic parenchymal remodeling and downstream extra-hepatic sequelae. Estimating CP grade from CECT provides an exploratory setting for examining how structural imaging phenotypes relate to clinically assessed hepatic reserve[3,4]. This task nevertheless presents an inherent structure–function semantic gap, motivating representation mappings that disentangle grade-relevant visual evidence while limiting reliance on spurious associations [5].

Conventional deep learning for medical imaging scans typically relies on slice-level 2D CNNs or volumetric 3D architectures. However, assigning patient-level CP grades to individual 2D slices introduces substantial label noise [3,4,6], while 3D CNNs may face a strict resolution-context trade-off, where memory-constrained downsampling obscures fine-grained parenchymal textures [7]. Motivated by computational pathology [8,9], we treat each CECT scan as a "bag" and its slices as "instances." This formulation matches patient-level CP supervision: it avoids assigning the bag label to every slice, accommodates variable-length scans, and selectively aggregates heterogeneous evidence. We therefore formulate CP grading as a weakly supervised MIL task and compare representative MIL architectures. However, applying conventional MIL to multi-source physiological decompensation exposes two incompatibilities.

Spatial feature entanglement. Clinically, intrinsic functional reserve is dictated by intra-hepatic parenchymal remodeling [10], whereas extra-hepatic signs (e.g., ascites, splenomegaly) are downstream systemic sequelae [11]. Conventional spatially-agnostic MIL encoders tend to confound these signals, capturing spurious co-occurrences rather than discrete morphological origins.

Associative shortcut learning. Most MIL models compress a bag into a single global embedding $\mathbf{z}$ and learn an unconstrained K-way classifier $p(y \mid \mathbf{z})$. This treats stages as independent labels, lacks semantic anchoring to stage-wise incremental evidence, and encourages shortcut features that correlate with labels but are not stage-specific pathology [12]. For progressive, cumulative decompensation, this entangles baseline and incremental cues in $\mathbf{z}$ and degrades generalization under domain shift.

We propose SLICE-MIL (Fig. 1), a spatial-logical framework with two phases: *structural isolation* and *semantic anchoring*. ADDA decouples intra- and extra-hepatic signals, then orthogonal filtration and residual routing separate the compensated basis S_0 (CP-A phenotype and grade-irrelevant artifacts) from pathological increments $(\mathbf{E}_1, \mathbf{E}_2)$. Since unconstrained routing is semantically underdetermined (Table 2), we add semantic anchoring via representation-level coun-

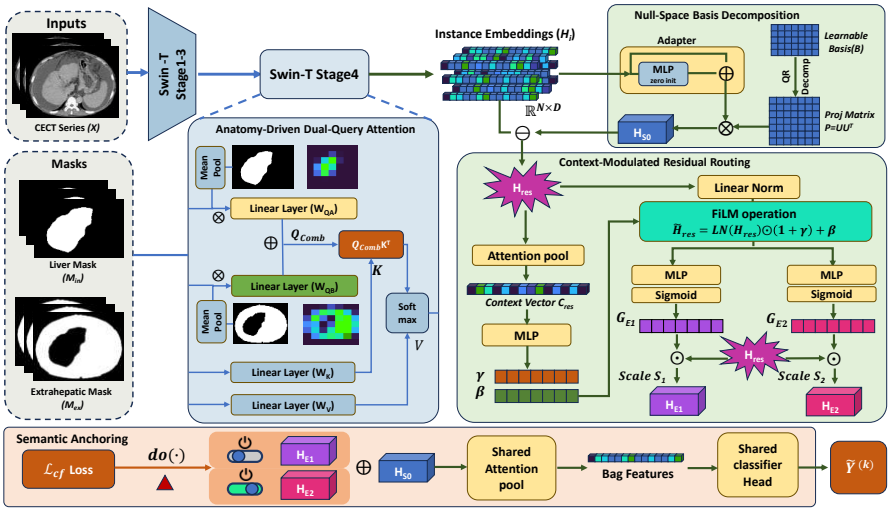

Fig. 1. SLICE-MIL. *ADDA* decouples intra-/extra-hepatic queries. Orthogonal decomposition and routing isolate the compensated(CP-A) basis (S_0) from pathological increments ($\mathbf{E}_1, \mathbf{E}_2$). *Semantic Anchoring* applies counterfactual ablations to regularize dominance-ordered staging logic.

terfactual ablations, enforcing a dominance order $E_2 \succ E_1 \succ S_0$ consistent with decompensation pathophysiology [1].

Our contributions are:

- **Structural isolation:** We propose ADDA to separate intra/extra-hepatic features via anatomy-gated queries, reducing cross-regional confounding.
- **Semantic anchoring:** We introduce representation-level counterfactual ablations to enforce dominance-ordered semantics, yielding an ordinal regularizer for MIL.
- **Clinical utility:** SLICE-MIL provides a robust framework for CECT-based Child-Pugh estimation and achieves the best performance among the evaluated MIL baselines.

2 Methodology

SLICE-MIL sequentially disentangles spatial and ordinal evidence. ADDA separates heterogeneous regions, while orthogonal filtration, residual routing, and representation-level interventions organize stage-specific increments.

2.1 Spatial Disentanglement: Anatomy-Driven Dual-Query Attention

Standard self-attention projects all tokens into a single unconstrained query subspace ($\mathbf{Q}=\mathbf{X}\mathbf{W}_Q$), potentially entangling intra-/extra-hepatic distributions and

promoting spurious cross-regional co-occurrences[13]. To address this, we replace standard self-attention in Swin Stage 4 with Anatomy-Driven Dual-Query Attention (ADDA).

Given the intermediate sequence $\mathbf{X} \in \mathbb{R}^{N\times D}$, the liver and extrahepatic masks are adaptively average-pooled to the Stage-4 feature resolution of 7×7, yielding soft regional masks $\mathbf{M}_{in}, \mathbf{M}_{ex} \in [0,1]^{N\times 1}$. Used only as coarse regional guidance for query construction, these masks reduce dependence on precise anatomical boundary delineation. The two regions are projected through separate query transformations:

$$\mathbf{Q}_{Comb} = (\mathbf{X} \odot \mathbf{M}_{in})\mathbf{W}_{QA} + (\mathbf{X} \odot \mathbf{M}_{ex})\mathbf{W}_{QB} \tag{1}$$

Under globally shared key-value projections ($\mathbf{K} = \mathbf{X}\mathbf{W}_K$, $\mathbf{V} = \mathbf{X}\mathbf{W}_V$), the attention output is $\mathrm{Softmax}(\mathbf{Q}_{Comb}\mathbf{K}^\top/\sqrt{d_k})\mathbf{V}$. Mathematically, the disjoint masks render Eq. 1 a piecewise affine transformation, projecting heterogeneous regions into distinct query subspaces. Retrieving from a shared dictionary $(\mathbf{K}, \mathbf{V})$ acts as a structural regularization, mitigating query distribution collapse and constraining the instance embeddings $\mathbf{H}_i$ to preserve spatially decoupled signals for the subsequent logical routing.

2.2 Logical Disentanglement: Orthogonal Filtration and Residual Routing

While ADDA isolates the spatial origins of pathological signals, we formalize their ordinal contributions to staging via an explicit latent decomposition (Fig. 2). It summarizes the computational dependencies among the disentangled latent components and should not be interpreted as a causal DAG. The instance embeddings $\mathbf{H}_i$ are decomposed into three functionally isolated components: a CP-A basis $\mathbf{H}_{S0}$ (CP-A phenotype), an early evidence pool $\mathbf{H}_{E1}$ (pathology driving CP-A→B), and a severe evidence pool $\mathbf{H}_{E2}$ (pathology driving CP-B→C).

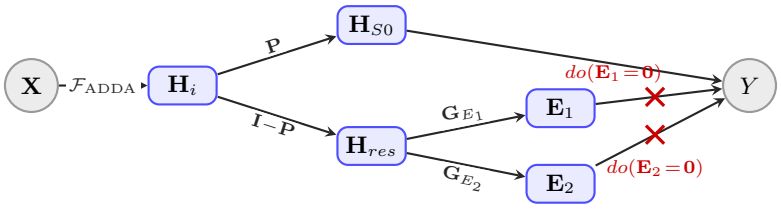

Fig. 2. Logical Dependency Graph of SLICE-MIL. ADDA produces spatial embeddings $\mathbf{H}_i$. The projector $\mathbf{P}$ isolates the basis $\mathbf{H}_{S0}$, while routing gates $\mathbf{G}_{E_j}$ extract ordinal evidence components. Cross marks denotes a representation-level *do*-operation implemented by ablating the selected component.

Null-Space Basis Decomposition (NSBD). To isolate the endogenous basis, we define a learnable matrix $\mathbf{B} \in \mathbb{R}^{D\times r}$. Performing a reduced QR decomposition

($\mathbf{B}=\mathbf{UR}$) yields an orthonormal basis $\mathbf{U} \in \mathbb{R}^{D\times r}$ such that $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_r$. We then construct a rank-r orthogonal projector $\mathbf{P}=\mathbf{U}\mathbf{U}^\top$ [14]. The spatial embeddings $\mathbf{F}$, stabilized by a zero-initialized residual Adapter, are decomposed into the basis subspace and its orthogonal complement:

$$\mathbf{H}_{S0} = \mathbf{F}\mathbf{P}, \quad \mathbf{H}_{res} = \mathbf{F}(\mathbf{I}-\mathbf{P}) \tag{2}$$

This structural idempotency mathematically guarantees token-wise orthogonality: $\langle \mathbf{h}_{S0,n}, \mathbf{h}_{res,n}\rangle = 0, \forall n$. By projecting $\mathbf{H}_{S0}$ into the null-space of pathological increments, we explicitly preclude information leakage, forcing residuals to capture only stage-specific variance.

Context-Modulated Residual Routing (CMRR). Local pathological tokens are inherently myopic; disambiguating early versus severe decompensation requires macroscopic context. We introduce a global-to-local feedback loop to dynamically route the residual manifold $\mathbf{H}_{res}$ into stage-specific evidence pools.

First, we distill the holistic residual context $\mathbf{C}_{res} \in \mathbb{R}^{1\times D}$ via Gated Attention Pooling [15]:

$$\mathbf{C}_{res} = \sum_{n=1}^{N} a_n \mathbf{h}_{res,n}, \quad a_n = \frac{\exp(\mathbf{w}^\top \mathbf{m}_n)}{\sum_{j=1}^{N}\exp(\mathbf{w}^\top \mathbf{m}_j)} \tag{3}$$

where $\mathbf{m}_n = \tanh(\mathbf{V}_1\mathbf{h}_{res,n}) \odot \sigma(\mathbf{V}_2\mathbf{h}_{res,n})$ suppresses non-informative tokens ($\mathbf{V}_1, \mathbf{V}_2 \in \mathbb{R}^{L\times D}$, $\mathbf{w} \in \mathbb{R}^{L\times 1}$).

Next, we regress affine parameters $[\boldsymbol{\gamma};\boldsymbol{\beta}] = \mathrm{MLP}(\mathbf{C}_{res})$ to dynamically recalibrate token-level residuals via FiLM [16]: $\tilde{\mathbf{H}}_{res} = \mathrm{LN}(\mathbf{H}_{res}) \odot (1+\boldsymbol{\gamma}) + \boldsymbol{\beta}$. Continuous gating matrices $\mathbf{G}_{Ej} = \sigma(\mathrm{MLP}_j(\tilde{\mathbf{H}}_{res}))$ then execute dense soft-routing into additive increments:

$$\mathbf{H}_{Ej} = (\tilde{\mathbf{H}}_{res} \odot \mathbf{G}_{Ej})\boldsymbol{\Sigma}_j, \quad j \in \{1,2\} \tag{4}$$

with channel-wise capacity calibration $\boldsymbol{\Sigma}_j = \mathrm{diag}(\mathbf{S}_j)$. Rather than imposing rigid algebraic priors, we rely on gradient-driven intervention objectives to supervise this functional disentanglement.

2.3 Semantic Anchoring via Counterfactual Intervention

While NSBD ensures structural separation, the routing remains semantically agnostic. Inspired by the intervention of structural causal models(SCM) [17,18], we introduce representation-level *do*-operations on the disentangled evidence components. Unlike an SCM, our framework does not specify observable causal variables, structural equations, or a causal DAG. Specifically, $do(\mathbf{E}_j = \mathbf{0})$ denotes replacing a selected evidence component with zero before global pooling. These operations construct counterfactual latent representations and provide clinically guided ordinal regularization.

The isolated components are additively fused, allowing the attention mechanism to dynamically recalibrate over the active evidence manifold. Formally, we generate counterfactual bag features $\mathbf{z}_{cf}^{(k)}$ by zeroing specific evidence pools, subject to logically shifted targets $\tilde{Y}^{(k)}$:

- **Case 1** ($do(\mathbf{E}_2 = \mathbf{0})$): $\mathbf{z}_{cf}^{(1)} = \text{AttnPool}(\mathbf{H}_{S0} + \mathbf{H}_{E1})$, $\tilde{Y}^{(1)} = \min(y, 1)$. Mathematically caps progression at CP-B.
- **Case 2** ($do(\mathbf{E}_1 = \mathbf{0})$): $\mathbf{z}_{cf}^{(2)} = \text{AttnPool}(\mathbf{H}_{S0} + \mathbf{H}_{E2})$, $\tilde{Y}^{(2)} = y \cdot \mathbb{I}(y \neq 1)$. Enforces dominance-ordering: $\mathbf{E}_1$ is strictly necessary for early progression ($y=1$), yet $\mathbf{E}_2$ remains independently sufficient to override into CP-C ($y=2$).
- **Case 3** ($do(\mathbf{E}_1 \wedge \mathbf{E}_2 = \mathbf{0})$): $\mathbf{z}_{cf}^{(3)} = \text{AttnPool}(\mathbf{H}_{S0})$, $\tilde{Y}^{(3)} = 0$. Strictly anchors the null-space col($\mathbf{P}$) to the CP-A basis.

Optimization Objective. Let $\mathcal{L}_{CE}$ denote the class-weighted cross-entropy. Operating over the factual state $\mathbf{z}_{fact} = \text{AttnPool}(\mathbf{H}_{S0} + \mathbf{H}_{E1} + \mathbf{H}_{E2})$ and counterfactual variants, the joint objective is:

$$\mathcal{L}_{total} = \mathcal{L}_{CE}\big(\text{Head}(\mathbf{z}_{fact}), y\big) + \lambda_{cf} \sum_{k=1}^{3} \mathcal{L}_{CE}\big(\text{Head}(\mathbf{z}_{cf}^{(k)}), \tilde{Y}^{(k)}\big) \tag{5}$$

This explicitly coerces the routing gates to strip redundant correlations from $\mathbf{H}_{res}$, ensuring $\mathbf{E}_1$ and $\mathbf{E}_2$ encode genuinely stage-specific, ordinal functional increments.

3 Experiments

3.1 Dataset and Implementation Details

We retrospectively curated a multi-center cohort of 767 portal venous phase[19] CECT scans. [5]The exclusion criteria were applied to reduce major factors that could confound CP assessment: (1) severe imaging artifacts; (2) history of massive hepatectomy, liver transplantation, or TIPS; (3) concurrent non-hepatic comorbidities independently interfering with CP indicators; (4) missing serological records. Liver parenchyma and effective body contour masks were extracted via TotalSegmentator [20] and verified by expert radiologists.

The cohort is institutionally partitioned into an internal discovery set ($N = 566$; CP-A/B/C = 241/211/114; 5-fold CV) and a strictly isolated external test set ($N = 201$; CP-A/B/C = 55/102/44). External metrics report mean±std across five models from internal CV. We employ a Swin-Tiny backbone [21] initialized with ImageNet-1K weights, optimized end-to-end via AdamW for 100 epochs. The counterfactual weight is $\lambda_{cf} = 0.5$. Quadratic Weighted Kappa (QWK) serves as the primary metric, rigorously penalizing non-adjacent misclassifications.

[5] This retrospective study was approved by the Ethics Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (No. TJ-IRB202407064). The requirement for written informed consent was waived because only de-identified data were used.

Table 1. Comparison with representative MIL methods(mean±std). **Bold**/underline: best/2nd. Methods without an asterisk utilize the ResNet-50 backbone. Methods denoted with an asterisk (*) utilize our proposed ADDA-Swin-T backbone.

Method	Internal (CV)			External Test		
	QWK↑	AUC↑	Acc↑	QWK↑	AUC↑	Acc↑
ABMIL [15]	.652±.119	.780±.074	.650±.093	.449±.099	.731±.059	.531±.049
CLAM [22]	.628±.122	.773±.073	.661±.059	.392±.158	.699±.100	.513±.078
TransMIL [8]	.656±.049	.804±.036	.661±.070	.487±.034	.760±.015	.516±.049
ACMIL [9]	.602±.181	.759±.097	.630±.163	.427±.128	.721±.082	.506±.066
MambaMIL [23]	.611±.050	.767±.045	.643±.063	.490±.037	.754±.023	.557±.052
SLICE-MIL	.774±.021	.873±.029	.744±.028	.603±.040	.796±.041	.626±.018
ABMIL*	.806±.019	.872±.014	.767±.013	.586±.045	.813±.013	.610±.036
CLAM*	.815±.011	.888±.012	.772±.017	.638±.024	.829±.011	.637±.015
TransMIL*	.810±.018	.880±.007	.769±.022	.592±.050	.823±.016	.612±.032
ACMIL*	.818±.016	.891±.012	.785±.025	.579±.062	.830±.019	.628±.041
MambaMIL*	.813±.018	.883±.010	.779±.020	.610±.029	.826±.003	.616±.026
SLICE-MIL*	**.835±.009**	**.916±.008**	**.799±.014**	**.683±.040**	**.859±.005**	**.687±.019**

Table 2. Ablation of key components. Baseline employs Swin-T as the feature encoder and ABMIL as the aggregator. Blank cells indicate the use of baseline components. All metrics are higher-better.

Model	ADDA	Route	$\mathcal{L}_{cf}$	Internal (CV)			External		
				QWK	AUC	Acc	QWK	AUC	Acc
Baseline				.761±03	.854±04	.736±05	.554±04	.759±07	.561±06
+ADDA	✓			.806±02	.872±01	.767±01	.586±05	.813±01	.610±04
+Route		✓		.795±02	.876±02	.753±02	.569±09	.773±01	.576±02
+Anchor		✓	✓	.812±02	.897±01	.781±03	.638±03	.834±02	.648±03
Full	✓	✓	✓	**.835±01**	**.916±01**	**.799±01**	**.683±04**	**.859±01**	**.687±02**

3.2 Results and Ablation

Table 1 shows substantial external performance degradation for several spatially agnostic MIL methods(e.g., CLAM's QWK drops by 37.6%). This cross-domain gap is consistent with sensitivity to source-specific variation and motivates spatially differentiated representation learning. Among the evaluated methods, SLICE-MIL achieves the highest external QWK of 0.683. The ADDA-Swin-T variants also exhibit consistent external improvements across different MIL aggregators. Together with the controlled ablation in Table 2, these results support the use of region-specific query encoding before global aggregation.

Table 2 shows complementary contributions from structural separation and semantic constraints. Routing without $\mathcal{L}_{cf}$ yields relatively high external variability ($\sigma = 0.088$), whereas adding $\mathcal{L}_{cf}$ reduces variability and improves performance, supporting its role as an ordinal regularizer. The full model achieves the

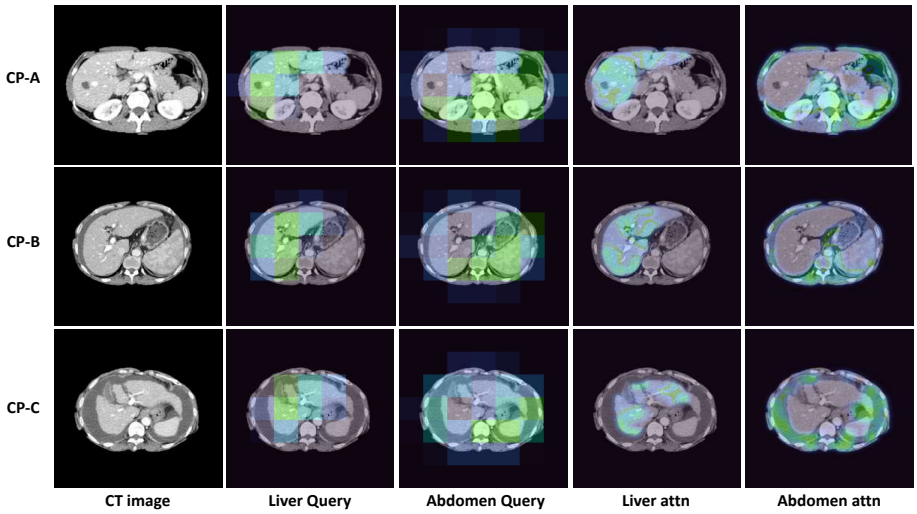

Fig. 3. Spatial activation heatmaps (rows: CP-A/B/C; cols: CT, Liver Query, Abdomen Query, Liver Attn, Abdomen Attn).

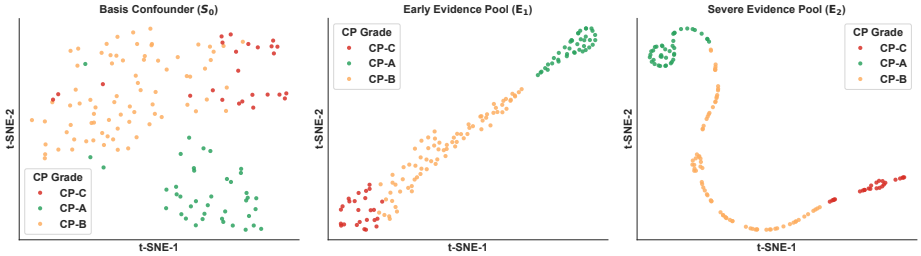

Fig. 4. t-SNE of disentangled subspaces (external set).

best internal and external results, indicating the joint benefit of spatial guidance, residual routing, and semantic anchoring.

3.3 Visualization and Interpretability

Anatomical Disentanglement. Fig. 3corroborates our anatomy-driven dual-query design: the Liver Query emphasizes intra-hepatic remodeling, whereas the Abdomen Query highlights extra-hepatic signs such as ascites, suggesting reduced spatial mixing before MIL aggregation.

Topological Alignment with Disease Logic. Figure 4 shows substantial grade overlap in $\mathbf{H}_{S0}$, gradual ordering in $\mathbf{E}_1$, and clearer separation of compensated and decompensated cases in $\mathbf{E}_2$. These patterns are consistent with the intended ordinal semantics of the learned components.

4 Conclusion

SLICE-MIL combines anatomy-driven attention and counterfactual anchoring for CECT-based Child-Pugh grading, improving cross-domain performance among the evaluated MIL methods. Limitations include the lack of broader 2.5D/3D comparisons, untested robustness to systematic segmentation failures. CECT-only estimation is intended as complementary imaging evidence and cannot replace standard multi-source CP assessment. Future work will use longitudinal multimodal data to investigate causal relationships between biochemical indicators and structural imaging phenotypes.

Acknowledgments. This study was partly supported by the Joint Funds of the National Natural Science Foundation of China (U23A201325 to Xiaoping Chen), the Joint Funds of the National Natural Science Foundation of Hubei (2025AFD842 to Qi Cheng), Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0525400, 2024ZD0525406 to Qi Cheng), and the Youth Science and Technology Talent Support Program of the Guizhou Association for Science and Technology (GASTYESS202535 to Yunsong Peng).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. European Association for the Study of the Liver: EASL clinical practice guidelines: Management of hepatocellular carcinoma. J. Hepatol. **69**(1), 182–236 (2018). https://doi.org/10.1016/j.jhep.2018.03.019
2. Pugh, R.N.H., Murray-Lyon, I.M., Dawson, J.L., Pietroni, M.C., Williams, R.: Transection of the oesophagus for bleeding oesophageal varices. Br. J. Surg. **60**(8), 646–649 (1973). https://doi.org/10.1002/bjs.1800600817
3. Thüring, J., Rippel, O., Haarburger, C., Merhof, D., Schad, P., Bruners, P., Kuhl, C.K., Truhn, D.: Multiphase CT-based prediction of Child-Pugh classification: A machine learning approach. Eur. Radiol. Exp. **4**(1), 20 (2020). https://doi.org/10.1186/s41747-020-00148-3
4. Huang, Z., Zhang, G., Liu, J., Huang, M., Zhong, L., Shu, J.: LRFNet: A deep learning model for the assessment of liver reserve function based on Child-Pugh score and CT image. Comput. Methods Programs Biomed. **223**, 106993 (2022). https://doi.org/10.1016/j.cmpb.2022.106993
5. Castro, D.C., Walker, I., Glocker, B.: Causality matters in medical imaging. Nat. Commun. **11**, 3673 (2020). https://doi.org/10.1038/s41467-020-17478-w
6. Campanella, G., et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. **25**(8), 1301–1309 (2019). https://doi.org/10.1038/s41591-019-0508-1
7. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016). https://doi.org/10.1109/3DV.2016.79

8. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., Zhang, Y.: TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. Adv. Neural Inf. Process. Syst. **34**, 2136–2147 (2021). https://proceedings.neurips.cc/paper/2021/hash/10c272d06794d3e5785d5e7c5356e9ff-Abstract.html
9. Zhang, Y., Li, H., Sun, Y., Zheng, S., Zhu, C., Yang, L.: Attention-challenging multiple instance learning for whole slide image classification. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) Computer Vision – ECCV 2024. LNCS, vol. 15111, pp. 125–143. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-73668-1_8
10. Smith, A.D., et al.: Liver surface nodularity quantification from routine CT images as a biomarker for detection and evaluation of cirrhosis. Radiology **280**(3), 771–781 (2016). https://doi.org/10.1148/radiol.2016151542
11. Zardi, E.M., Di Matteo, F.M., Pacella, C.M., Sanyal, A.J.: Invasive and non-invasive techniques for detecting portal hypertension and predicting variceal bleeding in cirrhosis: A review. Ann. Med. **46**(1), 8–17 (2014). https://doi.org/10.3109/07853890.2013.857831
12. Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., Wichmann, F.A.: Shortcut learning in deep neural networks. Nat. Mach. Intell. **2**(11), 665–673 (2020). https://doi.org/10.1038/s42256-020-00257-z
13. Wang, T., Zhou, C., Sun, Q., Zhang, H.: Causal attention for unbiased visual recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3091–3100 (2021). https://openaccess.thecvf.com/content/ICCV2021/html/Wang_Causal_Attention_for_Unbiased_Visual_Recognition_ICCV_2021_paper.html
14. Ravfogel, S., Elazar, Y., Gonen, H., Twiton, M., Goldberg, Y.: Null it out: Guarding protected attributes by iterative nullspace projection. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7237–7256. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.acl-main.647
15. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. In: Dy, J., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136. PMLR (2018). https://proceedings.mlr.press/v80/ilse18a.html
16. Perez, E., Strub, F., de Vries, H., Dumoulin, V., Courville, A.: FiLM: Visual reasoning with a general conditioning layer. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, pp. 3942–3951. AAAI Press (2018). https://doi.org/10.1609/aaai.v32i1.11671
17. Pearl, J.: Causality: Models, Reasoning, and Inference. 2nd edn. Cambridge University Press, Cambridge (2009). https://doi.org/10.1017/CBO9780511803161
18. Pawlowski, N., Coelho de Castro, D., Glocker, B.: Deep structural causal models for tractable counterfactual inference. Adv. Neural Inf. Process. Syst. **33**, 857–869 (2020). https://proceedings.neurips.cc/paper/2020/hash/0987b8b338d6c90bbedd8631bc499221-Abstract.html
19. Heo, S., Lee, S.S., Choi, S.H., Kim, D.W., Park, H.J., Kim, S.Y., Lee, S.J., Kim, K.M., Shin, Y.M.: CT rule-in and rule-out criteria for clinically significant portal hypertension in chronic liver disease. Radiology **309**(1), e231208 (2023). https://doi.org/10.1148/radiol.231208
20. Wasserthal, J., et al.: TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. Radiol. Artif. Intell. **5**(5), e230024 (2023). https://doi.org/10.1148/ryai.230024

21. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9992–10002. IEEE (2021). https://doi.org/10.1109/ICCV48922.2021.00986
22. Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nat. Biomed. Eng. **5**(6), 555–570 (2021). https://doi.org/10.1038/s41551-020-00682-w
23. Yang, S., Wang, Y., Chen, H.: MambaMIL: Enhancing long sequence modeling with sequence reordering in computational pathology. In: Linguraru, M.G., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. LNCS, vol. 15004, pp. 296–306. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72083-3_28