nature portfolio

Peer Review File

Deep Learning CT Signature for Predicting Early Liver Metastases in Pancreatic Ductal Adenocarcinoma

Corresponding Author: Professor Shenghong Ju

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
This study evaluated the utility of Mamba-based predictive model integrating imaging features from pancreatic tumor and the liver to determine the risk of occult liver metastases in a multi-institutional cohort. The model showed good performance in risk stratification and in predicting patient survival. In a subgroup of patients, the radiotranscriptomic analysis provided plausible biologic explanation of the model.

Major strengths:
1. Multi-center study with 1063 patients from 5 institutions, and the CTs were obtained on a variety of vendors, which can potentially increase the generalizability to other sites.
2. Analysis of imaging features from both primary pancreatic tumor and the liver is relatively novel.
3. Identification of patients at risk of early liver recurrence and patients most likely to benefit from NAT are clinically important problems.
4. Inclusion of the radiotranscriptomic analysis to provide plausible biologic explanation of the model.

Major weaknesses:
1. The definition of liver metastases within 6 months postop as “occult liver metastasis” is not a commonly accepted definition. Occult liver metastasis usually refers to detection of liver metastasis during surgical exploration that occur soon (within 1 month) of the CT. 6 months is a long window of opportunity for the liver to develop metastases. Consider using other terms such as “early liver recurrence” or “high risk for liver metastases” to convey the idea that the patients are at high risk, but the liver metastases were not necessarily present on the CT.
2. The rate of “occult liver metastases” ~28% (within 6 months) seem high, based on published literature and the experience at my institution. Median recurrence for upfront surgery is ~12-18 months. Therefore, this model may not be applicable at other sites.
3. The patients who underwent upfront surgery vs. neoadjuvant therapy where not randomized, which may introduce a source of bias, despite the propensity analysis.

Other comments:
1. Introduction. Statement of “Over half of these hepatic recurrences occur within 6 months” from References 5, 9. These references did not seem to state that. Ref 5 cites a Cochrane review describing staging laparoscopy detecting unresectable disease following CT (the usual definition for occult metastatic disease). Please clarify.
2. Introduction. Neoadjuvant therapy is meant to target patients with high-risk disease, not just patients with subclinical liver disease.
3. Results. Please add preoperative local staging (resectable, borderline resection, unresectable) and progression free survival and overall survival for the whole cohort and subgroups in the text or tables. The local staging is especially relevant for the patients from Center 5 who underwent upfront surgery or NAT.
4. Results P.9 line 1. Typo in “before or after mating”. Should be “matching”.
5. Methods. Liver segmentation. Did the segmentation exclude benign lesions (e.g., cysts, hemangiomas) that may affect the imaging features of the liver?
6. Table 1, extended Table 1-3: “Resection region” should be changed to “Resection margin”.
7. The proportion of patients who did not undergo adjuvant treatment seemed high in the internal cohorts (23.8-25.7%), and that may have an important impact on survival.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
In this manuscript, Zhao et al. present a carefully designed study that applies the emerging Mamba architecture to predict occult liver metastasis (OLM) in pancreatic ductal adenocarcinoma (PDAC) using multi-phase CT. The proposed dual-region model, jointly analyzing the primary tumor and liver parenchyma, is biologically well-grounded as it reflects the metastatic cascade. The robust performance across multiple cohorts, together with the potential to guide neoadjuvant therapy (NAT) decisions, underscores the strong translational value of this work. This is an interesting and timely study that addresses a clinically important question with a novel AI architecture, biologically rational modeling, and strong validation. With minor clarifications and revisions as outlined below, the manuscript will be further strengthened. Overall, this paper is suitable for publication in Nature Communications pending minor revisions.

Specific Comments:

1. Choice of Mamba architecture: Please expand on the rationale for adopting the Mamba framework rather than conventional transformer- or CNN-based models. A clear explanation will highlight the novelty and technical advantages of this choice.
2. Spatial feature representation: Flattening 3D features into 1D sequences may risk losing spatial context. Could the authors clarify how Mamba's modeling capability effectively preserves and leverages the spatial information in CECT images?
3. Biological interpretability of dual branches: The two-branch design (tumor vs. liver) is conceptually elegant and biologically rational. It would strengthen the paper to discuss whether the model identifies interpretable and biologically meaningful patterns—for example, which types of features from the tumor and parenchyma branches may drive predictive performance.
4. Clinical heterogeneity across cohorts: Since NCCN guidelines recommend NAT for high-risk features such as borderline resectable disease or elevated CA19-9, it would be helpful to address potential clinicopathological heterogeneity between the development and application cohorts.
5. Multi-modality integration: The integration of transcriptomic and radiomic findings is a commendable strength of this work. Could the authors comment on why a more formal multi-modality integration approach was not undertaken?
6. Fusion mechanism clarity: In Figure 1, the fusion mechanism of the two-branch network is not clearly illustrated. Please revise the figure to explicitly describe how tumor and liver features are concatenated or integrated at the decision level.
7. Figure 4 labeling: The Kaplan-Meier survival curves are valuable, but the labeling of axes should be clarified to improve interpretability.

(Remarks on code availability)
The Github repository provides detailed information about the proposed model, including Instructions about preprocessing, training, and inference.

Reviewer #5

(Remarks to the Author)
In this manuscript, Zhao et. al. presents a Mamba-based deep learning model that integrates imaging features from both the pancreatic tumor and the liver to predict occult liver micrometastases (OLM) in PDAC. The model demonstrates robust performance (AUC 0.806–0.890) in a large, multi-institutional cohort and provides clinically actionable stratification, as only high-risk patients derived significant survival benefit from neoadjuvant therapy. Radiotranscriptomic validation further supports the biological relevance of the risk groups, strengthening the translational impact of the study. Overall, the work represents a clinically significant advance by offering a noninvasive and generalizable tool for guiding NAT decision-making in PDAC. Its originality and strong clinical applicability make it highly valuable for the field, and the paper is suitable for acceptance after several minor amendments.
The work demonstrates several key strengths:
1. Novelty: the study applied a Mamba-based deep learning framework to jointly analyze pancreatic tumor and liver imaging for OLM prediction. The dual-organ imaging approach is conceptually innovative and clinically logical, considering that the liver is the most common metastatic site in PDAC.
2. Clinical and Biological Relevance: The model shows strong predictive performance in a large, multi-institutional cohort, underscoring its accuracy and generalizability. Notably, it identifies patients who benefit from neoadjuvant therapy (NAT), addressing a critical gap in PDAC management. Radiotranscriptomic validation further confirms the biological aggressiveness of the high-risk group, enhancing the translational and clinical relevance of the finding.
3. Methodology and Statistical Rigor: The methodology is sound, using a large multi-institutional dataset and appropriate statistical analyses (AUC, hazard ratios, propensity score matching). The results support the conclusions, with no major flaws, though more detail on model architecture, training, preprocessing, and cross-validation would improve reproducibility.
Overall, this is an important and clinically impactful study that could change the way PDAC patients are stratified for NAT. I recommend acceptance after minor revision, mainly to further strengthen the presentation. Specifically, the authors should:
1. Provide additional clarification on how Mamba compares to other deep learning architectures (e.g., CNNs, Transformers)

for this application in the Introduction.
2. Include a more detailed discussion of potential implementation challenges in real-world clinical practice.
3. Rephase the abstract using present tense and avoid grammar mistake.
4. Rearrange Fig.1 by shifting the clinical information in 1e to 1a. The layout for original Fig.1a and Extended Fig.1 is redundant.
5. It is suggested to arrange the subtitle in Result to “Overall study design” rather than “clinicopathological characteristics” for a better summarization.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
This study developed and validated a model to predict early liver metastases in patients who underwent surgical resection of pancreatic cancer. These results are clinically impactful in triaging patients most likely to benefit from surgery. The authors have adequately addressed the reviewers' comments regarding the clinical and machine learning aspects.

Minor comments:
In Extended Data Table 1 and 2: "Differential" should be "Differentiation". There is also a typo in "Lymphvascular invasion" should be "Lymphovascular invasion".

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The authors have made substantial and high-quality revisions to the manuscript. The revised version demonstrates clear improvements in methodological rigor, clarity of presentation, and clinical relevance. The responses are thorough, logically articulated, and well supported by additional analyses. I have no further comments.

(Remarks on code availability)
The source codes are fine for research reproducibility.

Reviewer #5

(Remarks to the Author)

I appreciate the authors’ careful and constructive revisions made in response to the reviewers’ comments. The updated manuscript has addressed all points raised in the review, including:
1. Additional clarification in the Introduction regarding Mamba architecture and its comparison to other deep learning approaches (CNNs, Transformers).
2. Expanded discussion on the potential challenges and considerations for real-world clinical implementation.
3. Revision of the abstract to use present tense with improved grammar and clarity.
4. Reorganization of Figure 1 by placing the clinical information from Fig. 1e into Fig. 1a, removing redundancy between the original Fig. 1a and Extended Fig. 1.
5. Adjustment of the Results section subtitle to “Overall study design” for better representation of the content.
The manuscript is now clearer, more informative, and easier to follow. The scientific conclusions are well supported by the data, and the study continues to demonstrate strong clinical significance and potential translational impact. All requested revisions have been satisfactorily completed. I recommend the manuscript for publication in its current form.

(Remarks on code availability)

Reviewer #6

(Remarks to the Author)
This manuscript addresses an important clinical challenge in pancreatic ductal adenocarcinoma, namely the high incidence of early liver metastases following apparently curative surgery. By integrating multi-phase CT features from both the primary tumor and the liver, the authors propose a deep learning framework to predict early liver metastases and explore whether the resulting risk stratification can identify patients who benefit from neoadjuvant therapy (NAT).

The conceptual motivation for incorporating liver imaging features is sound and distinguishes this work from many prior radiomics and deep learning studies that focus exclusively on the primary tumor. The multi-center design, inclusion of an external validation cohort, and the attempt to assess calibration and clinical utility represent clear strengths. Nevertheless, several major issues require clarification or additional analysis before the conclusions can be fully supported.

Major concerns

1. Interpretation of early liver metastases (ELM)
The study defines ELM as the occurrence of liver metastasis within six months after surgery. While this endpoint is clinically meaningful, it is not equivalent to direct evidence of pre-existing occult liver micrometastases at the time of imaging. As currently presented, it remains unclear whether the model is identifying true preoperative micrometastatic disease or a broader phenotype of biologically aggressive tumors that recur early after surgery.

This distinction is important for interpreting the clinical implications of the model. Additional sensitivity analyses using alternative time windows (e.g., shorter or longer postoperative intervals) and a more explicit discussion of this limitation would help clarify what biological process the model is most likely capturing.

2. Robustness of the neoadjuvant therapy (NAT) benefit analysis
The observation that NAT appears to confer overall survival benefit only in the model-defined high-risk group is intriguing but represents the most methodologically vulnerable aspect of the study. Although propensity score matching is performed, the matching variables appear limited and may not fully capture key clinical factors influencing NAT selection and treatment completion.

Moreover, the retrospective design raises concerns about potential selection bias and immortal time bias, particularly if patients who initiated NAT but did not proceed to surgery were excluded from the analysis. Clarification of patient inclusion criteria and handling of such cases is necessary. To strengthen this claim, the authors should consider expanding the set of covariates used for matching, reporting balance diagnostics, and formally testing treatment–risk interaction effects in a regression framework.

3. Independent contribution of liver-derived imaging features
A central premise of the manuscript is that liver imaging features provide complementary information beyond the primary tumor. While this is supported conceptually by the proposed architecture, more systematic evidence is needed to demonstrate their independent contribution.

In particular, clearer ablation analyses comparing tumor-only, liver-only, and combined tumor–liver models would be helpful. Given the multi-center nature of the dataset, it is also important to exclude the possibility that liver features are inadvertently capturing center-specific imaging characteristics rather than biologically meaningful signals.

4. Generalizability of the model
Although an external validation cohort is included, all centers are drawn from a relatively homogeneous geographic and ethnic population. This limitation should be more clearly acknowledged when discussing potential clinical deployment. Additional subgroup analyses stratified by scanner vendor or imaging parameters would further support claims of robustness.

Minor concerns

1. More detail on the quality control of automated segmentations would improve transparency, including any assessment of inter-reader variability during manual correction.

2. The risk threshold is defined using the Youden index; reporting additional clinically interpretable operating points (e.g., fixed sensitivity or negative predictive value) would enhance translational relevance.

3. Calibration performance is mentioned but could be more fully quantified, particularly in the external validation cohort.

4. The radiotranscriptomic analysis is based on a relatively small sample size and should be framed more explicitly as exploratory.

5. Additional details on preprocessing and inference pipelines would further improve reproducibility.

Summary

In summary, this is a well-motivated and technically solid study that addresses a clinically relevant problem in PDAC. However, several key claims—particularly regarding the biological interpretation of early liver metastases and the selective benefit of neoadjuvant therapy—require additional clarification and supporting analyses. Addressing the major concerns outlined above would substantially strengthen the manuscript.

(Remarks on code availability)
The authors provide a publicly accessible GitHub repository with source code and a README describing the overall pipeline. The repository is generally well organized and facilitates understanding of the proposed architecture.

However, full reproducibility is currently limited by the absence of example data, pretrained weights, and detailed instructions for reproducing the reported experiments across cohorts. Clarifying the preprocessing steps, model configuration files, and inference workflow would further enhance the utility of the code as a community resource.

Version 2:

Reviewer comments:

Reviewer #6

(Remarks to the Author)
I have no further comments.

(Remarks on code availability)

Response to Reviewer #3:

Overall Response:

We sincerely thank the reviewer for the positive evaluation and high recommendation of our study. We greatly appreciate the thoughtful and detailed feedback regarding the definition of occult liver metastasis and various aspects of manuscript presentation. In response, we have carefully addressed all comments in a point-by-point manner. The manuscript has been substantially improved through this rigorous review process, and we are grateful for the reviewer's insightful suggestions and meticulous reading of our work.

Comment #1:

The definition of liver metastases within 6 months postop as "occult liver metastasis" is not a

commonly accepted definition. Occult liver metastasis usually refers to detection of liver metastasis during surgical exploration that occur soon (within 1 month) of the CT. 6 months is a long window of opportunity for the liver to develop metastases. Consider using other terms such as "early liver recurrence" or "high risk for liver metastases" to convey the idea that the patients are at high risk, but the liver metastases were not necessarily present on the CT.

Response:

Thank you for your insightful comment. We agree that the term "occult liver metastasis" lacks a universally standardized definition. To improve clarity and avoid misinterpretation, we have revised the terminology throughout the manuscript. As suggested, we now use the term "early liver recurrence" to describe metastases detected within 6 months postoperatively. This better reflects the clinical scenario.

Comment #2:

The rate of "occult liver metastases" ~28% (within 6 months) seem high, based on published literature and the experience at my institution. Median recurrence for upfront surgery is ~12-18 months. Therefore, this model may not be applicable at other sites.

Response:

Thank you for your positive comments and questions. Indeed, the definition of early liver metastases and reported rates differ in literatures, ranging from approximately 20% within 6 months in some studies to as high as 45% within 3 months in others[1,2]. These discrepancies across institutions are likely attributable to differences in imaging protocols, surveillance strategies, and clinical practices.

The rate of "early liver metastases" ~28%, which seems slightly higher. We provide the

following potential explanations for the finding: many patients with PDAC, especially those with ELM, often present with early symptoms or biochemical abnormalities, they tend to undergo postoperative imaging studies early and repeatedly. This practice may increase the observed proportion of metastatic disease in our data. Conversely, individuals who are asymptomatic or lost to systematic follow-up may be under-represented because timely imaging and regular surveillance are lacking, and partial of these patients without regular follow-up may be excluded. Consequently, our reported rate of liver metastasis may be probably over-estimated. We also explicitly acknowledge this selection bias as a limitation of the present study: *"Second, the patient population was derived exclusively from East Asian centers, and the relatively high incidence of ELM in our population may limit the generalizability offindings to other ethnic or geographic populations with different clinical characteristics."*

Reference

1. Wang, F. *et al.* Evaluation of occult liver metastases in pancreatic adenocarcinoma by diffusion-weighted related magnetic resonance imaging. *J Magn Reson Imaging* **62**, 536–548 (2025).

2. Liu, J.-B. *et al.* Development and validation of a predictive model for occult liver metastasis in pancreatic ductal adenocarcinoma using subjective imaging and clinical data. *Cancer Med* **14**, e71280 (2025).

Comment #3:

The patients who underwent upfront surgery vs. neoadjuvant therapy where not randomized, which may introduce a source of bias, despite the propensity analysis.

Response:

We sincerely thank the reviewer for this insightful comment. Although our dataset represents one

of the largest multi-institutional cohorts with detailed imaging, treatment, and outcome data. We fully acknowledge that the non-randomized nature of treatment allocation in this retrospective study may introduce potential selection bias, even after rigorous propensity score matching. As suggested, we have explicitly addressed this limitation in the revised manuscript: "*First, despite the use of propensity score matching and multivariable adjustments, the retrospective study designinherently introduces potential selection bias and unmeasured confounding,which may influence the observed associations.*" Looking forward, we agree that higher-level evidence is needed to confirm the clinical utility of our model in guiding treatment selection. As such, we have also added a statement: "*Finally, our findings require prospective validation in large and multicenter cohorts. Future studies should further explore the utility of this model in guiding other treatment strategies, such as adjuvant therapy intensity or immunotherapy selection, and assess its real-time clinical integration in multidisciplinary treatment workflows.*"

Comment #4:

Introduction. Statement of "Over half of these hepatic recurrences occur within 6 months" from References 5, 9. These references did not seem to state that. Ref 5 cites a Cochrane review describing staging laparoscopy detecting unresectable disease following CT (the usual definition for occult metastatic disease). Please clarify.

Response:

We sincerely appreciate your attention to detail and the opportunity to improve the clarity and accuracy of our manuscript. As you mentioned, these 2 reviews state PDAC is an aggressive disease, typically with early systemic spread, and most patients eventually develop systemic disease recurrence, suggesting the presence of systemic micrometastases. These references do not

specifically support the claim that "over half of hepatic recurrences occur within six months postoperatively." We acknowledge that the original citation was therefore inappropriate, and we apologize for this inaccuracy.

To better reflect the current understanding and to enhance the precision of our statement, we have revised the text as follows: "*Yet, disease recurrence is distressingly common, affecting up to 80% of patients. The liver is the most frequent site of distant recurrence, with a substantial proportion of liver recurrences occurring in the early postoperative period, often within six months after curative-intent resection [6–8]. These early liver metastases (ELM) are often characterized by aggressive biological behavior, likely reflecting the presence of occult micrometastatic disease, and are associated with dismal post-recurrence survival, typically less than six months [6,9,10].*" We hope the revised statement can more accurately capture the clinical reality of early liver metastasis in PDAC, align with the cited references, and improve the contextual coherence of our discussion.

Reference

6. Zhang, X.-P. *et al.* Early and late recurrence patterns of pancreatic ductal adenocarcinoma after pancreaticoduodenectomy: a multicenter study. *Int J Surg* **109**, 785–793 (2023).

7. Groot, V. P. *et al.* Defining and Predicting Early Recurrence in 957 Patients With Resected Pancreatic Ductal Adenocarcinoma. *Ann. Surg.* **269**, 1154–1162 (2019).

8. Groot, V. P. *et al.* Patterns, timing, and predictors of recurrence following pancreatectomy for pancreatic ductal adenocarcinoma. *Ann Surg* **267**, 936–945 (2018).

9. Groot, V. P. *et al.* Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. *Annals of Surgery* **267**, 936–945 (2018).

10. Tsilimigras, D. I. *et al.* Liver metastases. *Nat Rev Dis Primers* **7**, 27 (2021).

Comment #5:

2. Introduction. Neoadjuvant therapy is meant to target patients with high-risk disease, not just patients with subclinical liver disease.

Response:

You are absolutely correct. Neoadjuvant therapy (NAT) is increasingly used to treat patients with high-risk features beyond just subclinical liver involvement—such as borderline resectability, elevated CA19-9, large primary tumors, or suspicious lymphadenopathy.We have revised the Introduction to reflect this broader rationale: "*Neoadjuvant therapy (NAT) represents a promising therapeutic approach for high-risk PDAC patients, particularly those likely with high risk of occult micrometastatic disease.*"

Comment #6:

2. Results. Please add preoperative local staging (resectable, borderline resection, unresectable) and progression free survival and overall survival for the whole cohort and subgroups in the text or tables. The local staging is especially relevant for the patients from Center 5 who underwent upfront surgery or NAT.

Response:

We sincerely appreciate the reviewer's insightful comment. As suggested, we have supplemented the preoperative local staging (resectable, borderline resectable), as well as progression-free survival (PFS) and overall survival (OS) for the entire cohort and relevant subgroups, which are now clearly presented in Table 1.

Table 1 | The baseline clinicopathological characteristics of patients in training and validation cohorts

Characteristics	Training cohort (n = 450)	Internal validation cohort (n = 113)	External validation cohort (n = 342)
Age (years)[a]	64.73 ± 9.31	63.34 ± 9.98	64.79 ± 9.35
Sex			
Female	173 (38.4)	48 (42.5)	146 (42.7)
Male	277 (61.6)	65 (57.5)	196 (57.3)
Tumor location			
Head or neck	308 (68.4)	79 (69.9)	199 (58.2)
Body or tail	142 (31.6)	34 (30.1)	143 (41.8)
Local staging			
Resectable	392 (87.1)	94 (83.1)	299 (87.4)
Borderline resectable	58 (12.9)	19 (16.8)	43 (12.6)
Tumor size (cm)[b]	3.00 (2.30, 3.50)	3.00 (2.50, 3.80)	3.00 (2.40, 3.70)
CA19-9 (U/mL)			
≤ 300	258 (57.3)	63 (55.8)	171 (50.0)
> 300	192 (42.7)	50 (44.2)	171 (50.0)
Total bilirubin (μmol/L)[b]	30.96 (12.60, 150.20)	24.80 (12.40, 119.00)	15.00 (9.70, 84.20)
Direct bilirubin (μmol/L)[b]	15.70 (4.22 , 112.70)	11.60 (4.10, 89.00)	5.50 (3.80, 68.80)
Albumin (g/L)[b] Globulin	39.15 (33.40, 43.70)	38.00 (31.60, 43.30)	43.30 (39.72, 46.10)
(g/L)[b] Differentiation	28.30 (25.10, 33.40)	28.80 (24.50, 34.20)	25.80 (23.60, 28.20)
Well	241 (53.6)	56 (49.6)	159 (46.5)
Poor	209 (46.4)	57 (50.4)	183 (53.5)
T stage			
T1	79 (17.6)	15 (13.3)	80 (23.4)
T2	284 (63.1)	76 (67.3)	213 (62.3)
T3 or T4	87 (19.3)	22 (19.5)	49 (14.3)
TNM stage			
≤ IIa	295 (65.6)	70 (61.9)	188 (55.0)
≥ IIb	155 (34.4)	43 (38.1)	154 (45.0)
Nerve infiltration	365 (81.1)	85 (75.2)	278 (81.3)
Lymphovascular invasion	133 (29.6)	34 (30.1)	112 (32.7)
Resection margin			
R0	429 (95.3)	109 (96.5)	321 (93.9)
R1	21 (4.7)	4 (3.5)	21 (6.1)
LNM	155 (34.4)	43 (38.1)	154 (45.0)
ELM	129 (28.7)	32 (28.3)	98 (28.7)
Adjuvant treatment			
Yes	219 (48.7)	60 (53.1)	236 (69.0)
No	107 (23.8)	29 (25.7)	36 (10.5)
Unknown	124 (27.6)	24 (21.2)	70 (20.5)
PFS	11.1 (9.3, 12.3)	11.2 (8.6, 14.8)	10.7 9.4, 12.8)
OS	23.2 (21.6, 27.3)	24.0 (19.1, 30.7)	28.0 (24.4, 30.9)

Except where indicated, data are numbers of patients, with percentages in parentheses.

CA19-9, carbohydrate antigen 19-9; R0, negative surgical margin; R1, positive surgical margin; LNM, lymph node metastasis; ELM, early liver metastases; PFS, progression free survival; OS, overall survival.

[a] mean ± standard deviation, [b] medians (interquartile range)

While all enrolled patients were classified as having localized disease (including both resectable and borderline resectable categories) at baseline based on multidisciplinary assessment, we fully agree with the reviewer that local staging is a critical determinant of clinical decision-making, particularly for patients from Center 5, who received either upfront surgery or neoadjuvant therapy (NAT). To address this, we have incorporated local staging as a covariate in Extended Data Table 4.

Extended Data Table 4 | The baseline preoperative clinical characteristics based on surgical approach, before and after propensity score matching

Characteristics	Total cohort			Prospensity score matched cohort		
	NAT + Surgery (n = 158)	Upfront Surgery (n = 238)	*P* value	NAT + Surgery (n = 123)	Upfront Surgery (n = 123)	*P* value
Age (years)			0.022			> 0.999
≤ 65	97 (61.4)	117 (49.2)		72 (60.6)	72 (58.5)	
> 65	61 (38.6)	121 (50.8)		51 (39.4)	51 (41.5)	
Sex			0.055			0.521
Female	80 (50.6)	96 (40.3)		51 (50.7)	57 (50.7)	
Male	78 (49.4)	142 (59.7)		72 (49.3)	66 (49.3)	
Tumor location			0.663			0.605
Head or neck	88 (55.7)	126 (52.9)		74 (59.2)	69 (59.2)	
Body or tail	70 (44.3)	112 (47.1)		49 (40.8)	54 (40.8)	
Local staging			0.164			0.262
Resectable	130 (82.3)	209 (87.8)		103 (83.7)	110 (89.4)	
Borderline resectable	28 (17.7)	29 (12.2)		20 (16.3)	13 (10.6)	
Tumor size (cm)			0.010			>0.999
≤ 3	62 (39.2)	126 (52.9)		52 (60.5)	52 (50.0)	
> 3	96 (60.8)	112 (47.1)		71 (39.5)	71 (50.0)	
CA 19-9 (U/mL)			0.378			0.898
≤ 300	75 (47.5)	125 (52.5)		57 (48.6)	56 (48.6)	
> 300	83 (52.5)	113 (47.5)		66 (51.4)	67 (51.4)	

Except where indicated, data are numbers of patients, with percentages in parentheses.

NAT, neoadjuvant therapy; CA 19-9, carbohydrate antigen 19-9; R0, negative surgical margin; R1, positive

surgical margin.

Besides, to minimize selection bias and confounding due to these imbalances in baseline prognostic factors, we performed propensity score matching (PSM) using key clinical variables, including age, sex, tumor size, CA19-9 level, and local staging. After matching, 123 patients remained in both the surgery-first and NAT groups. In this balanced cohort, we observed that model-defined high-risk patients who received NAT followed by surgery experienced significantly longer OS compared to those treated with upfront surgery (median OS: 17.4 vs. 34.1 months; HR: 1.87, 95% CI: 1.27–2.76, $P < 0.001$) (Fig. 4c). This survival benefit remained significant after PSM (median OS: 17.8 vs. 30.2 months; HR: 1.71, 95% CI: 1.04–2.80, $P = 0.019$) (Fig. 4d).

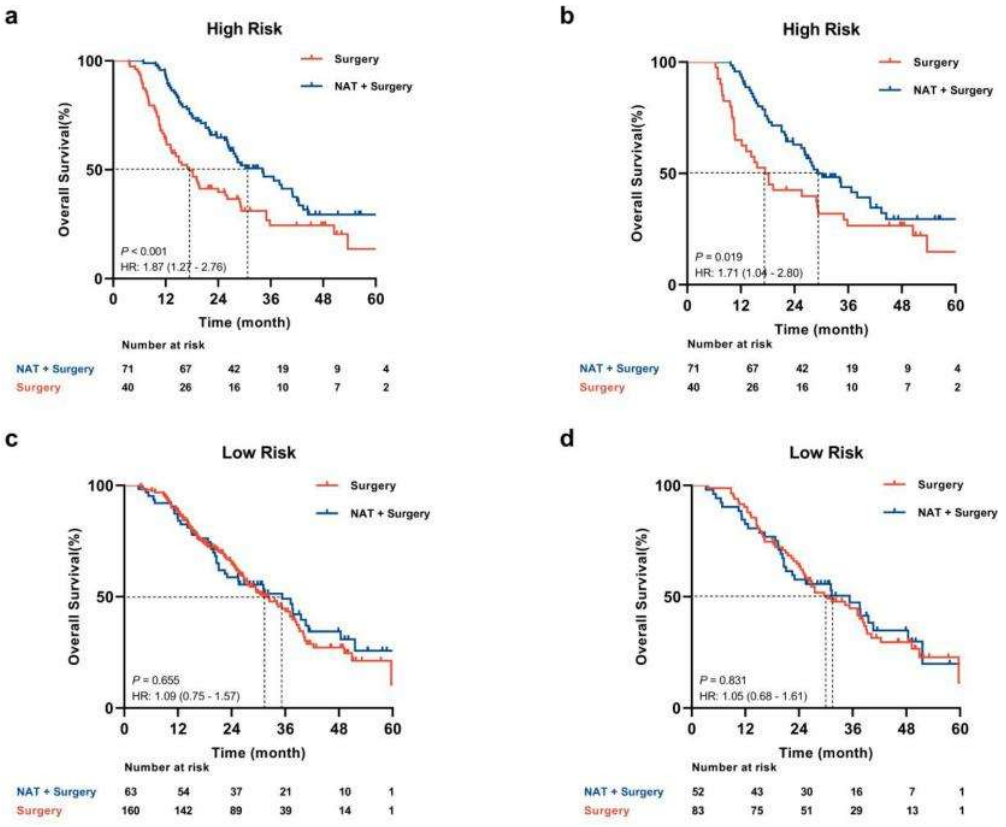


Fig. 4 | Mamba-based model predicts benefit from neoadjuvant chemotherapy. Kaplan-Meier curves of overall survival for patients who received or did not receive NAT in different risk groups

stratified by the Mamba-based model. The findings in the total cohort (a, c) versus the results in the cohort after propensity score matching (b, d). NAT, neoadjuvant therapy.

Comment #7:

4. Results P.9 line 1. Typo in “before or after mating”. Should be “matching”.

Response:

Thank you for identifying this typographical error. We apologize for the oversight. The text has been corrected from “before or after mating” to “*before or after matching*” in the revised manuscript.

Comment #8:

Methods. Liver segmentation. Did the segmentation exclude benign lesions (e.g., cysts, hemangiomas) that may affect the imaging features of the liver?

Response:

Thank you for your insightful comment regarding the impact of benign liver lesions on model’s performance. In our model, the liver was included as a key input via a dual-branch architecture, with whole-liver segmentation performed without exclusion of benign lesions (e.g., cysts, hemangiomas).

Our Mamba-based model is specifically designed to integrate spatiotemporal information from multiphase CT scans, enabling it to capture dynamic enhancement patterns across both the primary tumor and the liver. While benign hepatic lesions, predominantly cyst, hemangiomas and calcifications, do exhibit distinct enhancement kinetics, our framework effectively learns to differentiate these from malignancy-associated patterns through end-to-end training. Besides, we regard the presence of benign lesions as a natural perturbation that enhances model robustness in real-world settings.

To further address your concern, we conducted a supplementary analysis in a subset of patients with benign lesions manually excluded. This refinement did not yield a meaningful improvement in diagnostic performance, with AUCs of 0.882 (training), 0.803 (internal validation), and 0.780 (external validation), while substantially increasing segmentation burden. Given the minimal gain in accuracy and the added clinical impracticality, we recommend whole-liver segmentation as a robust and efficient strategy for model input. We appreciate this valuable comment, which prompted us to further validate our methodological choice.

Comment #9:

Table 1, extended Table 1-3: "Resection region" should be changed to "Resection margin". **Response:**

Thank you for your careful review and this accurate observation. We agree that "Resection region" was incorrect. The term has been revised to "*Resection margin*" in Table 1 and Extended Data Tables 1–3 to accurately reflect the pathological assessment.

Comment #10:

The proportion of patients who did not undergo adjuvant treatment seemed high in the internal cohorts (23.8-25.7%), and that may have an important impact on survival.

Response:

Thank you for your positive comments and questions. Thank you for this important observation. Indeed, the proportion of patients not receiving adjuvant therapy in the internal cohorts (23.8–25.7%) reflects real-world clinical heterogeneity, which arises from differences in postoperative complications, performance status, patient preference, and access to care.

PDAC is a highly aggressive and clinically heterogeneous disease, and treatment patterns

naturally vary across institutions. Our multicenter design captures this diversity, and we did not enforce treatment homogeneity across sites to preserve the real-world generalizability of our findings. Importantly, Survival analyses were performed within each cohort by first stratifying patients into high- and low-risk groups according to our model's output, and then comparing survival outcomes between these strata and we consistently observed significant survival differences favoring the low-risk group across all cohorts.

Response to Reviewer #4:

Overall Response:

Thank you very much for your thoughtful review and constructive feedback. We greatly appreciate your careful reading of our manuscript, your recognition of its scientific merit, and your valuable suggestions for improvement. Your insightful comments have helped us strengthen the quality and clarity of the work, and we have addressed all points in detail in the revised manuscript.

Comment #1:

Choice of Mamba architecture: Please expand on the rationale for adopting the Mamba framework rather than conventional transformer- or CNN-based models. A clear explanation will highlight the novelty and technical advantages of this choice.

Response:

Thank you for your insightful comment regarding the superiority of using Mamba. Mamba was selected for its ability to explicitly model spatial and temporal contrast variations through its selective state-space modeling (SSM) mechanism. Unlike CNNs that are limited by local receptive

fields or Transformers that incur quadratic complexity, Mamba employs a linear recurrent formulation to efficiently capture long-range voxel-level dependencies while maintaining scalability to high-resolution 3D data.

This design choice aligns closely with the clinical interpretive process, where radiologists evaluate temporal enhancement patterns across the liver and tumor over multiple phases. By recursively updating hidden states, Mamba achieves fine-grained spatiotemporal sensitivity and global contextual awareness that conventional architectures struggle to balance.

Furthermore, Mamba operates with linear computational complexity, a crucial advantage given our task involves handling large-scale data input--three high-resolution 3D CT volumes across a dual-branch architecture. By efficiently managing this substantial computational burden, Mamba proves exceptionally suited to our task, enabling our method to achieve significantly lower training and inference costs compared to Transformer-based models while maintaining modeling capacity.

Comment #2:

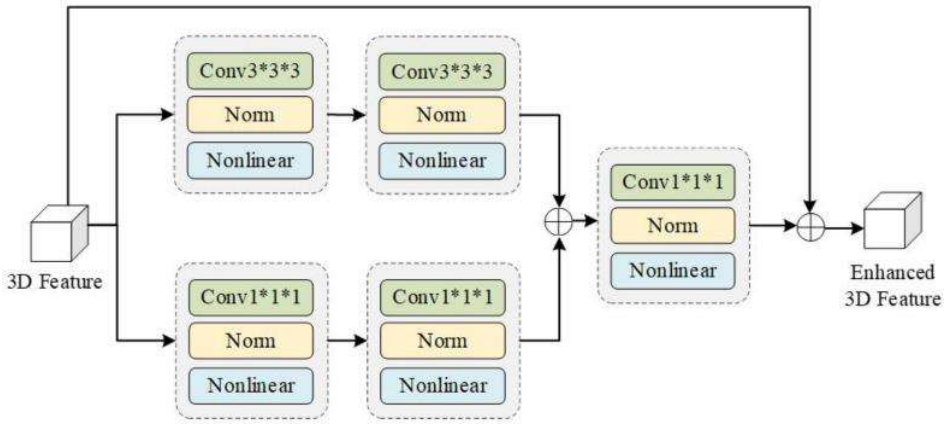
Spatial feature representation: Flattening 3D features into 1D sequences may risk losing spatial context. Could the authors clarify how Mamba's modeling capability effectively preserves and leverages the spatial information in CECT images?

Response:

Thank you for raising this important concern. We fully agree that flattening 3D features into a 1D sequence may risk losing spatial coherence. In our framework, this operation is carefully designed to enable fine-grained global modeling through dedicated spatial-scanning and temporal-scanning, while explicitly preserving spatial relationships through dedicated mechanisms.

Specifically, the flattening operation reorders multi-phase 3D features into a unified sequence with dimension (Batch Size, Channel, Slice × Width × Height), decoupling spatial and temporal contexts for separate but complementary modeling. This enables our Mamba-based model's spatial-scanning and temporal-scanning modules to capture fine-grained, phase-specific contrast-enhanced patterns across the liver and tumor region.

To preserve spatial context, the sequence is subsequently reshaped back into its original 3D lattice for further processing, ensuring structural consistency. Furthermore, to mitigate the loss of spatial information after flattening, we design a spatial-enhanced module (Supplement Figure 1) before each flattening operation that enriches the complementary details across different feature levels. Specifically, 3D features are processed through two parallel paths: One path employs two 3×3×3 convolution blocks to encode channel-wise spatial context. Each block consists of a convolution, a normalization layer, and a nonlinear activation layer. Another path applies two 1×1×1 convolution blocks to aggregate pixel-wise cross-channel context. Finally, the enhanced spatial features are obtained through residual connections. More details can be seen in the source code we provide.



Supplement Figure 1. Spatial-enhanced module in our framework that enriches the spatial information.

Comment #3:

Biological interpretability of dual branches: The two-branch design (tumor vs. liver) is conceptually elegant and biologically rational. It would strengthen the paper to discuss whether the model identifies interpretable and biologically meaningful patterns—for example, which types of features from the tumor and parenchyma branches may drive predictive performance.

Response:

We thank the reviewer for this thoughtful comment. We fully agree that biological interpretability is essential for clinical applicability. Our dual-branch design was motivated by the clinical rationale that tumor-related features (e.g., morphology, heterogeneity) and liver parenchyma features (e.g., texture, background liver condition) provide complementary biological cues for prediction. Therefore, we designed an attention-based tumor-liver fusion mechanism that adaptively balances the contribution of each branch through attention weighting. We summarize the fusion mechanism below and illustrate the challenge of direct visualization under this mechanism.

As shown in Supplement Figure 2, after the encoder stage, the three-phase features of each branch are concatenated to generate a pancreas tumor feature and a liver feature. Subsequently, each feature is refined through an intra-organ self-attention module to filter out noise and redundancy introduced by multiple downsamplings:

$$F' = \mathrm{SelfAttn}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where F’ denotes the enhanced tumor or liver feature, and Q, K, V represent the query, key, and value matrices generated from the tumor/liver features. The refined features are then processed with residual connections and layer normalization:

$$F'' = LN\ F' + F$$

where LN denotes the layer normalization function, F''represents the normalized tumor feature (F_t) or liver feature (F_o), and F denotes the original feature. Next, cross-attention is applied to enable interaction between the tumor and liver features. The feature of tumor branch generates matrices of Q_t' , K_t' , V_t' , while the liver branch generates matrices of Q_o' , K_o' , V_o'. The cross-attention mechanism is defined as:

$$F_t^{int} = \text{CrossAttn}\ \ Q'_t\ K'_o\ V'_o\ ,\ F_o^{int} = \text{CrossAttn}\ \ Q'_o\ K'_?\ V'_?\ ,$$

where F_t^{int} and F_o^{int} denote the interaction-enhanced tumor and liver features, respectively. These features are further refined by residual connections and layer normalization:

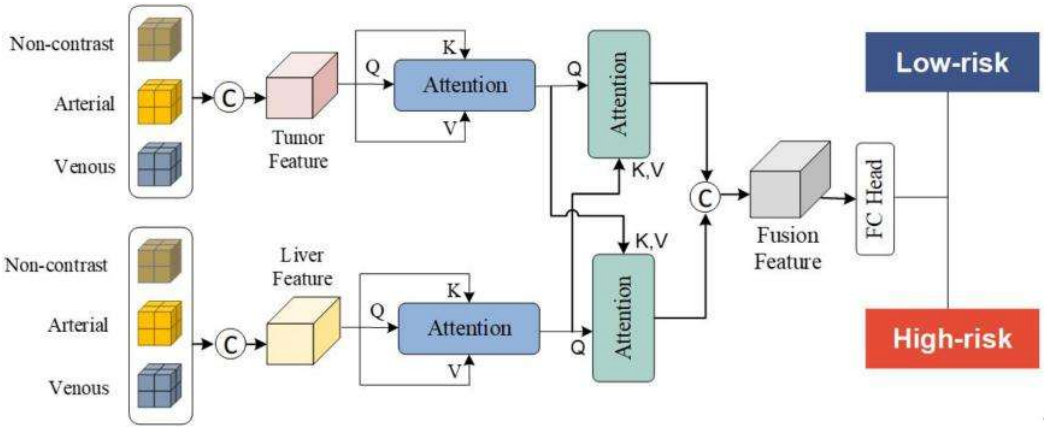
$$F''_t = LN\ F_t^{int} + F_t\ ,\ F''_o = LN\ F_o^{int} + F_o$$

Finally, the enhanced tumor and organ features are concatenated along the channel dimension to form the fused representation for final prediction. This design allows the model to leverage biologically complementary cues across the tumor and liver.

Experimental results demonstrate the effectiveness of this fusion mechanism. However, the designed fusion architecture poses inherent challenges for direct feature visualization (e.g., via Grad-CAM). First, both tumor and liver features are constructed by concatenating three-phase (non-contrast, arterial, and portal venous) representations. In subsequent processing, these three phases are modeled jointly as an integrated feature through intra-organ and cross-organ attention. In contrast, visualization requires phase-specific feature correspondence with the respective CT phases, which our fusion scheme does not explicitly preserve. Secondly, both the tumor and liver features are flattened into sequences and successively processed by intra-organ self-attention and cross-attention modules. These operations disrupt the original voxel-wise spatial correspondence,

making it difficult to project attention weights back to the 3D input domain. Furthermore, the incorporation of residual connections and layer normalization introduces both linear and nonlinear feature mixing, diffusing the contribution of individual neurons across the entire sequence. At the final decision stage, the tumor and liver representations are concatenated along the channel dimension, forming a fused high-dimensional embedding that cannot be directly disentangled into organ-specific spatial maps.

To address this challenge and enhance the biological plausibility of our model, we proactively incorporated complementary multi-omic analyses during the study design. Specifically, we performed radiotranscriptomic analysis by integrating imaging phenotypes with transcriptomic profiles from matched tumor tissues to reveal biologically meaningful associations.



Supplement Figure 2. The fusion mechanism of the pancreatic tumor branch and liver branch.

Comment #4:

Clinical heterogeneity across cohorts: Since NCCN guidelines recommend NAT for high-risk features such as borderline resectable disease or elevated CA19-9, it would be helpful to address potential clinicopathological heterogeneity between the development and application cohorts.

Response:

Thank you for this insightful comment. We agree that clinical heterogeneity between the development cohort and the NAT+surgery application cohort. To address this concern, we have presented a comprehensive comparison in Extended Data Table 3, which is described in the main text under the section "Assessment of the model's utility in guiding neoadjuvant therapy." This table highlights differences in key preoperative variables, including resectability status and CA19-9 levels, reflecting the NAT+surgery application cohort with higher-risk.

Also, to minimize potential bias arising from these imbalances, we subsequently performed propensity score matching (PSM) to balance the distribution of important prognostic factors between groups. We believe these additions strengthen the validity of our findings and provide a more nuanced understanding of the model's applicability in clinically heterogeneous populations.

Extended Data Table 3 | The clinical heterogeneity between the development and NAT+surgery cohorts

Characteristics	Training cohort (n = 450)	NAT + Surgery cohort (n = 158)	*P*
Age (years)			0.015
≤ 65	224 (49.8)	97 (61.4)	
> 65	226 (50.2)	61 (38.6)	
Sex			0.010
Female	173 (38.4)	80 (50.6)	
Male	277 (61.6)	78 (49.4)	
Tumor location			0.005
Head or neck	308 (68.4)	88 (52.9)	
Body or tail	142 (31.6)	70 (47.1)	
Local staging			0.068
Resectable	392 (81.7)	130 (82.3)	
Borderline resectable	58 (12.9)	28 (17.7)	
Tumor size (cm)			0.020
≤ 3	225 (50.0)	62 (39.2)	
> 3	225 (50.0)	96 (60.8)	
CA 19-9 (U/mL)			0.040
≤ 300	258 (57.3)	75 (47.5)	
> 300	192 (42.7)	83 (52.5)	

Except where indicated, data are numbers of patients, with percentages in parentheses.

CA 19-9, carbohydrate antigen 19-9; R0, negative surgical margin; R1, positive surgical margin.

Comment #5:

Multi-modality integration: The integration of transcriptomic and radiomic findings is a commendable strength of this work. Could the authors comment on why a more formal multi-modality integration approach was not undertaken?

Response:

We appreciate the reviewer's insightful comment regarding multi-modality integration. While formal fusion of multi-omics data is promising, our study focuses on a clinically urgent question: identifying patients with early liver metastases who may not benefit from upfront surgery. In clinical practice, preoperative imaging is the most accessible, rapid, and cost-effective tool for risk stratification, whereas transcriptomic and pathological data are typically available postoperatively. Therefore, our primary goal was to develop a preoperative predictive model using widely available radiomic features to guide clinical decision-making.

Nonetheless, we performed radiotranscriptomic analysis to provide biological interpretability to the imaging-derived signatures, thereby offering insights into the underlying tumor biology. It serves to bridge imaging phenotypes with molecular mechanisms. We agree with the reviewer that multi-modality integration holds great potential, and we will consider such approaches in future studies when addressing clinical questions where complementary preoperative data are available and could further improve predictive accuracy. Thank you again for this valuable suggestion.

Comment #6:

Fusion mechanism clarity: In Figure 1, the fusion mechanism of the two-branch network is not clearly illustrated. Please revise the figure to explicitly describe how tumor and liver features are concatenated or integrated at the decision level.

Response:

We thank the reviewer for this constructive comment. In response, we have revised **Figure 1** and further added more details in the Method parts of the revised manuscript to improve the clarity of the fusion mechanism in our dual-branch network. The updated figure extends the model schematic to more clearly illustrate the separate feature extraction pathways for tumor and liver regions, as well as their integration via cross-attention and concatenation at the decision level. We illustrate the fusion mechanism below.

After the encoder stage, the three-phase features of each branch are concatenated to generate a pancreas tumor feature and a liver feature. Subsequently, each feature is refined through an intra-organ self-attention module to filter out noise and redundancy introduced by multiple downsamplings:

$$F' = \mathrm{SelfAttn}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where F' denotes the enhanced tumor or liver feature, and Q, K, V represent the query, key, and value matrices generated from the tumor/liver features. The refined features are then processed with residual connections and layer normalization:

$$F'' = \mathrm{LN}(F' + F)$$

where LN denotes the layer normalization function, F''represents the normalized tumor feature (F_t) or liver feature (F_o), and F denotes the original feature. Next, cross-attention is applied to enable interaction between the tumor and liver features. The feature of tumor branch generates matrices of Q_t' , K_t' , V_t' , while the liver branch generates matrices of Q_o' , K_o' , V_o'. The cross-attention mechanism is defined as:

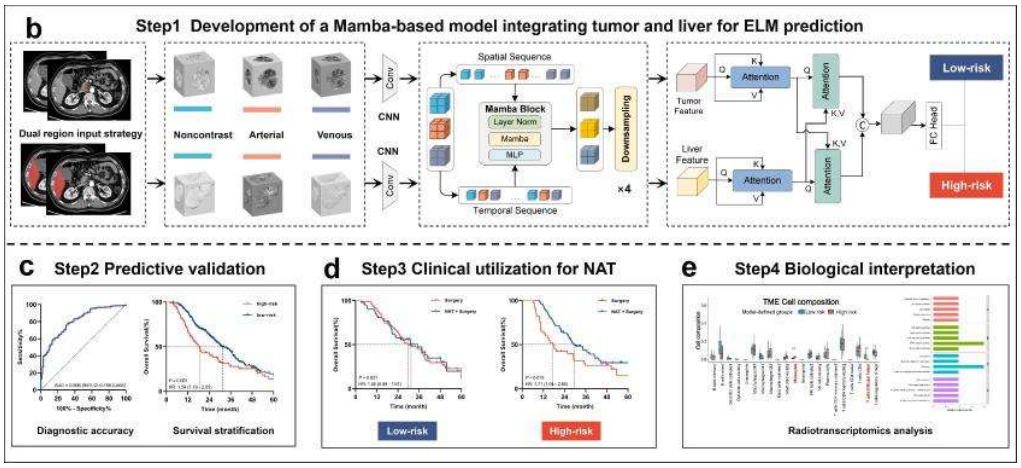
$$F_t^{int} = \mathrm{CrossAttn}(Q'_t, K'_o, V'_o) , F_o^{int} = \mathrm{CrossAttn}(Q'_o, K'_?, V'_?) ,$$

where F_t^{int} and F_o^{int} denote the interaction-enhanced tumor and organ features, respectively. These features are further refined by residual connections and layer normalization:

$$F''_t = LN\ F_t^{int} + F_t\ ,\ F''_o = LN\ F_o^{int} + F_o$$

Finally, the enhanced tumor and organ features are concatenated along the channel dimension to form the fused representation for final prediction.

This revision, combined with a reorganized layout that aligns with clinical workflow, enhances the overall interpretability of the model architecture. We appreciate the feedback, which has helped us better communicate the design and functionality of our framework.



Comment #7:

Figure 4 labeling: The Kaplan-Meier survival curves are valuable, but the labeling of axes should be clarified to improve interpretability.

Response:

We thank the reviewer for this constructive comment. We agree that the labeling in Figure 4 was insufficiently clear, particularly regarding the subpanel designations (a-d), which were inadvertently omitted in the previous version and may have caused confusion. In response, we have revised Figure 4 with improved axis labels, including explicit definitions of the survival

endpoints and time units, to enhance clarity and interpretability. Additionally, we have corrected labeling errors in the related panels of all Figures to ensure consistency and accuracy. We appreciate this valuable feedback.

Comment #8:

The Github repository provides detailed information about the proposed model, including Instructions about preprocessing, training, and inference.

Response:

We thank the reviewer for acknowledging the transparency and reproducibility of our work. Indeed, we have fully documented the model development pipeline in the public GitHub repository, including detailed instructions for data preprocessing, model training, and inference. We hope this facilitates both validation and future adaptation of our approach by the research community.

Response to Reviewer #5:

Overall Response:

Thank you very much for your positive comments and high recommendations on our study. This study evaluated the utility of Mamba-based predictive model integrating imaging features from pancreatic tumor and the liver to determine the risk of occult liver metastases in a multi-institutional cohort. The model showed good performance in risk stratification and in predicting patient survival. In a subgroup of patients, the radiotranscriptomic analysis provided plausible biologic explanation of the model.

Comment #1:

Provide additional clarification on how Mamba compares to other deep learning architectures (e.g., CNNs, Transformers) for this application in the Introduction.

Response:

Thank you for your insightful comment regarding the superiority of using Mamba. Mamba was selected for its ability to explicitly model spatial and temporal contrast variations through its selective state-space modeling (SSM) mechanism. Unlike CNNs that are limited by local receptive fields or Transformers that incur quadratic complexity, Mamba employs a linear recurrent formulation to efficiently capture long-range voxel-level dependencies while maintaining scalability to high-resolution 3D data.

This design choice aligns closely with the clinical interpretive process, where radiologists evaluate temporal enhancement patterns across the liver and tumor over multiple phases. By recursively updating hidden states, Mamba achieves fine-grained spatiotemporal sensitivity and global contextual awareness that conventional architectures struggle to balance.

Furthermore, Mamba operates with linear computational complexity, a crucial advantage given our task involves handling large-scale data input--three high-resolution 3D CT volumes across a dual-branch architecture. By efficiently managing this substantial computational burden, Mamba proves exceptionally suited to our task, enabling our method to achieve significantly lower training and inference costs compared to Transformer-based models while maintaining modeling capacity.

In response to the reviewer's suggestion, we have further clarified these advantages in the Introduction of the revised manuscript. Specifically, we now state: "*Mamba dynamically*

modulates information flow based on input content, enabling effective extraction of spatiotemporal features from sequential imaging data across diverse regions [26]." We have also added relevant references to support the use of Mamba in medical imaging to strengthen the contextual foundation of our methodological choice.

Reference

26. Qiu, B. *et al.* Multitask deep learning based on longitudinal CT images facilitates prediction of lymph node metastasis and survival in chemotherapy-treated gastric cancer. *Cancer Res* **85**, 2527–2536 (2025).

Comment #2:

Include a more detailed discussion of potential implementation challenges in real-world clinical practice.

Response:

Thank you for this valuable suggestion. We agree that a discussion of real-world implementation challenges is essential for translating predictive models into clinical practice. To better articulate both the clinical utility and translational limitations of our model, we have expanded the Discussion and Limitations sections accordingly.

With regard to clinical value, our findings highlight the potential of the imaging biomarker to support personalized management in pancreatic ductal adenocarcinoma (PDAC). For patients undergoing surgical resection: *Given that early recurrence remains a major contributor to poor outcomes in PDAC, our imaging biomarker demonstrates the potential to non-invasively identify high-risk patients who may benefit from intensified postoperative surveillance and tailored adjuvant therapy [2,22]. Therefore, These findings collectively demonstrate that our model has the potential to*

extract biologically relevant information from standard imaging phenotypes, thereby guiding clinical decision-making in real-world clinical practice [5,30,31]. In the neoadjuvant setting, current decision-making relies predominantly on anatomical criteria, which fail to capture underlying tumor biology: *Our findings suggest imaging-derived biological risk could refine patient selection for NAT, thereby addressing a key gap in current clinical guidelines and potentially informing future trial design based on biological risk rather than anatomical criteria alone* [33,34].

Nevertheless, this is a retrospective study, and findings require prospective validation in multicenter, diverse populations to assess generalizability and robustness: *Several limitations of this study warrant consideration. First, despite the use of propensity score matching and multivariable adjustments, the retrospective study designinherently introduces potential selection bias and unmeasured confounding,which may influence the observed associations. Second, the patient population was derived exclusively from East Asian centers, and the relatively high incidence of ELM in our populationmay limit the generalizability offindings to other ethnic or geographic populations with different clinical characteristics. Third, while our model showed promising results in predicting NAT benefit and stratifying prognosis, direct comparisons between patients who received chemotherapy and those who did not should be interpreted cautiously due to potential confounding. Finally, our findings require prospective validation in large and multicenter cohorts.*

Comment #3:

Rephase the abstract using present tense and avoid grammar mistake.

Response:

We sincerely thank the reviewer for the careful reading of our manuscript and for the valuable suggestion. In response, we have carefully revised the abstract to consistently use the present tense

and have corrected all grammatical errors. The revised version improves the clarity, accuracy, and overall readability of the abstract.

Comment #4:

Rearrange Fig.1 by shifting the clinical information in 1e to 1a. The layout for original Fig.1a and Extended Fig.1 is redundant.

Response:

We appreciate your valuable feedback on our manuscript. To avoid redundancy and improve clarity, we have revised Figure 1 to provide a clearer representation of the dual-branch Mamba model, now featuring an enhanced schematic that explicitly illustrates the multi-time-point inputs and cross-attention mechanism. The updated layout also eliminates overlap with Extended Data Figure 1, resulting in a more streamlined and cohesive presentation. Additionally, we have repositioned the clinical information from panel 1e to panel 1a, ensuring a more logical flow of information for readers. The revised Figure 1 is presented below. We believe these modifications we have made several refinements to Figure 1 to enhance its clarity and effectiveness in conveying our research findings. Thank you once again for your constructive comments.

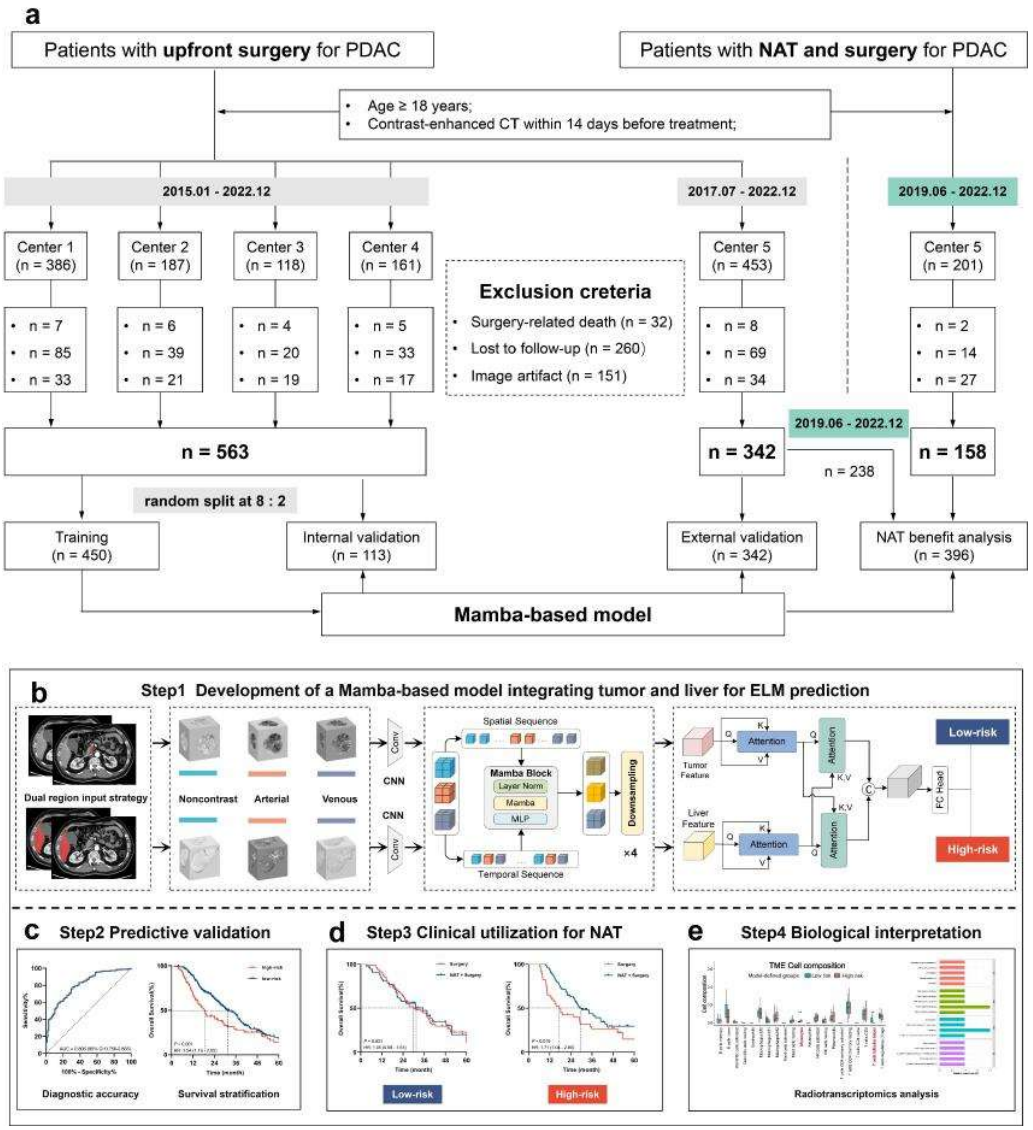


Fig. 1 | Overview of model's development, evaluation and clinical utilization. a, Patient inclusion and exclusion across different cohorts. **b,** The development of Mamba-based model for the prediction of ELM. **c,** Model evaluation. The performance of Mamba-based model in different cohorts. **d,** clinical utilization. We explore its pitential value in predicting benefit from NAT. **e,** Biological insights. Deciphering the underlying biological interpretation of the Mamba-based model. ELM, early liver metastases; PDAC, pancreatic ductal adenocarcinoma.

Comment #5:

It is suggested to arrange the subtitle in Result to "Overall study design" rather than "clinicopathological characteristics" for a better summarization.

Response:

We thank the reviewer for the valuable suggestion. We agree that "*Overall study design*" is more appropriate, and we have revised the subsection title accordingly. This change improves the clarity and organization of the Results section.

Dear Editor:

Re: “Deep Learning CT Signature for Predicting Early Liver Metastases and Neoadjuvant Therapy Benefit in Pancreatic Ductal Adenocarcinoma: A Multicenter Retrospective Study” (No.: NCOMMS-25-34821A).

We would like to extend our sincere gratitude to the reviewers for their thoughtful, insightful, and constructive comments on our manuscript. We have carefully considered all feedback and have revised the manuscript thoroughly in response to each point raised. All changes have been clearly marked in the revised manuscript using track changes.

Response to Reviewer #3:

Overall Response:

We sincerely thank the reviewer for the positive evaluation and high recommendation of our study.

Comment #1:

In Extended Data Table 1 and 2: "Differential" should be "Differentiation". There is also a typo in "Lymphvascular invasion" should be "Lymphovascular invasion".

Response:

Thank you for your insightful comment. We have corrected the spelling errors in the tables. "Differential" has been changed to "Differentiation", and "Lymphvascular invasion" has been changed to "Lymphovascular invasion" in both Extended Data Table 1 and Extended Data Table 2.

Response to Reviewer #4:

Overall Response:

Thank you very much for your thoughtful review; it has truly contributed to the quality of our manuscript.

Response to Reviewer #5:

Overall Response:

Thank you very much for your positive comments and high recommendations on our study.

Response to Reviewer #6:

Overall Response:

We sincerely thank you for taking the time to thoroughly review our manuscript and for providing such constructive and insightful comments. Your detailed feedback has been invaluable in identifying areas for improvement and enhancing the overall quality of our work. We have carefully considered each of your suggestions and concerns. In response, we have made significant revisions to the manuscript as outlined in our point-by-point response to the reviewers' comments. We believe that these changes have substantially improved the clarity, rigor, and impact of our study.

Comment #1:

Interpretation of early liver metastases (ELM)

The study defines ELM as the occurrence of liver metastasis within six months after surgery. While this endpoint is clinically meaningful, it is not equivalent to direct evidence of pre-existing occult liver micrometastases at the time of imaging. As currently presented, it remains unclear

whether the model is identifying true preoperative micrometastatic disease or a broader phenotype of biologically aggressive tumors that recur early after surgery.

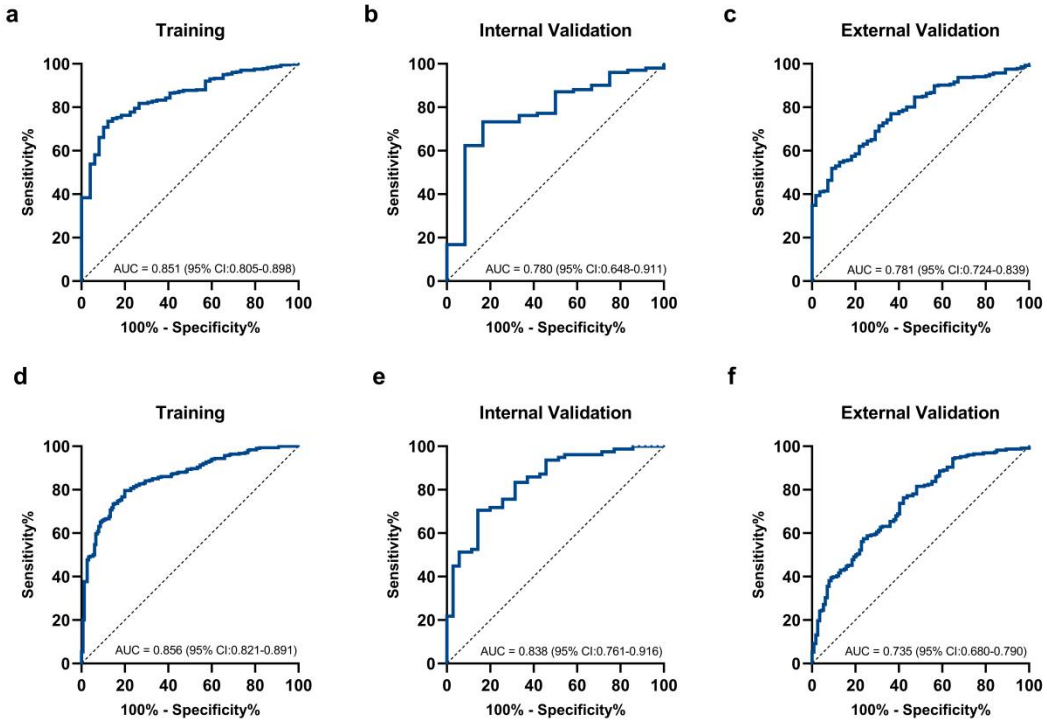
This distinction is important for interpreting the clinical implications of the model. Additional sensitivity analyses using alternative time windows (e.g., shorter or longer postoperative intervals) and a more explicit discussion of this limitation would help clarify what biological process the model is most likely capturing.

Response:

We appreciate the reviewer's valuable insight regarding the interpretation of Early Liver Metastases (ELM). We acknowledge that defining ELM as liver metastasis occurring within six months post-surgery, while clinically relevant as an outcome, does not provide direct evidence of pre-existing occult micrometastases at the time of preoperative imaging. It is indeed challenging to definitively distinguish between these micrometastases based solely on preoperative scans. Therefore, as suggested by reviewer #3, we have updated the terminology throughout the manuscript to consistently use "Early Liver Metastases (ELM)" instead of "occult liver metastases" to better reflect the nature of our endpoint.

As the reviewer mentioned, our deep learning model integrates features from both the primary pancreatic lesion and the liver parenchyma, potentially capturing the biological aggressiveness of the tumor effectively, thereby enabling accurate prediction of post-operative liver metastasis risk. Following the reviewer's suggestion, we conducted additional sensitivity analyses. We evaluated the model's performance in predicting liver metastases occurrence within shorter ($\leqq$ 3 months) and longer ($\leq$ 12 months) post-operative timeframes. These analyses showed consistent trends, with AUC values ranging from 0.735 to 0.856, supporting the model's ability to

stratify risk across different early recurrence windows. And we have added the result in the last of the part of Predictive value of the imaging biomarker for early liver metastases:“ *Beyond the primary endpoint, we also explored the model's predictive capability for liver metastasis occurrence across different time windows. Specifically, the model showed consistent predictive accuracy for both shorter-term (≤ 3 months post-surgery) and longer-term (≤ 12 months post-surgery) liver metastases development (Extended Data Fig. 6).* ”



Extended Data Fig. 6 | The performance of Mamba-based model to predict the liver metastases development in shorter-term and longer-term. (a-c) Predictive performance for short-term liver metastases (≤ 3 months). **(d-f)** Predictive performance for longer-term liver metastases (≤ 12 months)

Comment #2:

Robustness of the neoadjuvant therapy (NAT) benefit analysis

The observation that NAT appears to confer overall survival benefit only in the model-defined high-risk group is intriguing but represents the most methodologically vulnerable aspect of the

study. Although propensity score matching is performed, the matching variables appear limited and may not fully capture key clinical factors influencing NAT selection and treatment completion.

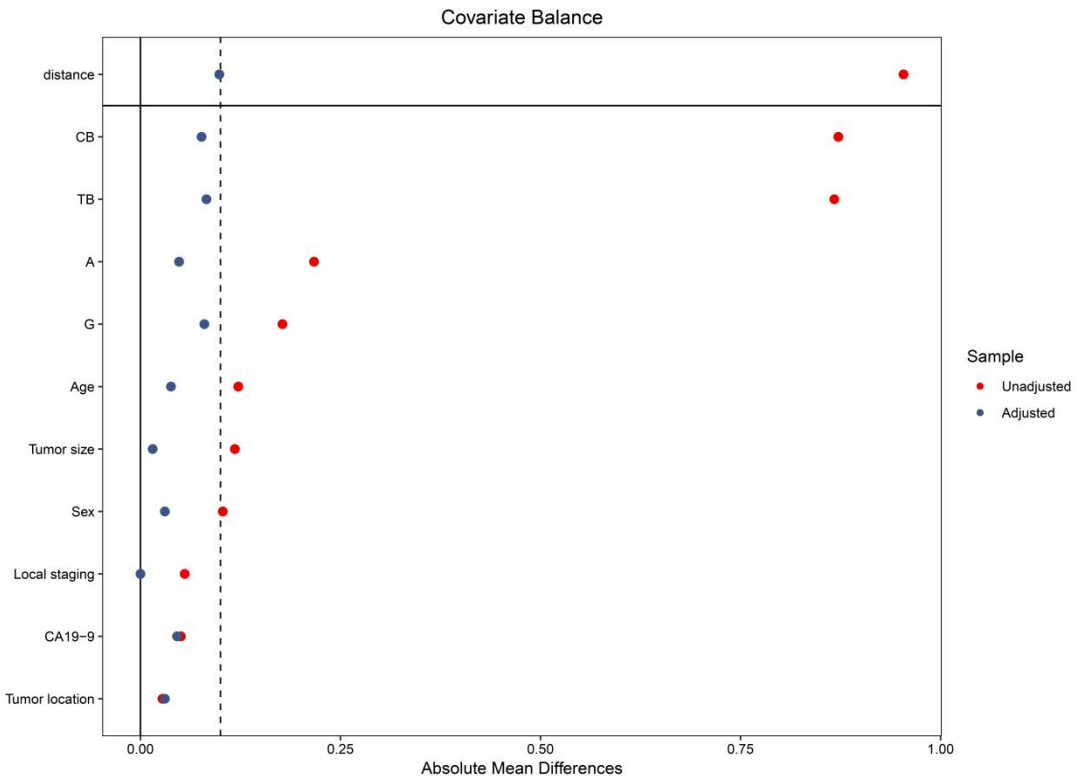
Moreover, the retrospective design raises concerns about potential selection bias and immortal time bias, particularly if patients who initiated NAT but did not proceed to surgery were excluded from the analysis. Clarification of patient inclusion criteria and handling of such cases is necessary. To strengthen this claim, the authors should consider expanding the set of covariates used for matching, reporting balance diagnostics, and formally testing treatment-risk interaction effects in a regression framework.

Reesponse：

Thank you for your insightful comments and interest in our exploratory finding. We appreciate your highly professional suggestions regarding the robustness of our analysis. As you correctly pointed out, this retrospective exploratory analysis faces inherent limitations, including potentially incomplete clinical information and unavoidable selection bias and immortal time bias due to its retrospective nature. As you known we enrolled patients who received NAT and surgery in our study, its hard for us to call back these patients who received NAT treatment but without surgery. We acknowledge these limitations and clearly state them in the Discussion section of the revised manuscript:“ *Third, while our model showed promising results in predicting NAT benefit and stratifying prognosis, we did not include patients who received NAT but did not proceed to surgical intervention. Direct comparisons between patients who received chemotherapy and those who did not should be interpreted cautiously due to potential confounding.*”

Also, in response to the reviewer’s suggestion to expand the set of covariates for matching,

we incorporated all available clinical parameters into PSM analysis. Following PSM with these variables, a balanced cohort of 262 patients (131 in each group) was successfully established. We have provided detailed reporting of the covariate balance diagnostics, including both graphical visualizations and table to demonstrate the effectiveness of the matching procedure. Furthermore, the results section of the main text has been updated to reflect these methodological improvements and the resulting balanced cohort characteristics:“ *Propensity score matching (PSM) was performed with potential clinical covariates. After matching, 131 patients were retained in each group with balanced baseline characteristics (Extended Data Table 4 and Extended Data Fig. 7).*”



Extended Data Fig. 7 | Covariate balance before and after propensity score matching. Absolute standardized mean differences for clinical variables comparing patients who received NAT followed by surgery versus those who underwent upfront surgery. NAT, neoadjuvant therapy

Extended Data Table 4 | The baseline preoperative clinical characteristics based on surgical approach, before and after propensity score matching

Characteristics	Total cohort			Prospensity score matched cohort		
	NAT + Surgery (n = 158)	Upfront Surgery (n = 238)	*P* value	NAT + Surgery (n = 131)	Upfront Surgery (n = 131)	*P* value
Age (years)			0.022			0.618
≤ 65	97 (61.4)	117 (49.2)		72 (55.0)	77 (58.8)	
> 65	61 (38.6)	121 (50.8)		59 (45.0)	54 (41.2)	
Sex			0.055			0.711
Female	80 (50.6)	96 (40.3)		67 (51.1)	63 (48.1)	
Male	78 (49.4)	142 (59.7)		64 (48.9)	68 (51.9)	
Tumor location			0.663			0.710
Head or neck	88 (55.7)	126 (52.9)		69 (52.7)	73 (55.7)	
Body or tail	70 (44.3)	112 (47.1)		62 (47.3)	58 (44.3)	
Local staging			0.164			> 0.999
Resectable	130 (82.3)	209 (87.8)		112 (85.5)	112 (85.5)	
Borderline resectable	28 (17.7)	29 (12.2)		19 (14.5)	19 (14.5)	
Total bilirubin (μmol/L)[b]	13.05 (9.00, 21.30)	15.10 (10.00, 61.60)	< 0.001	13.20 (9.10, 20.83)	12.30 (9.13, 21.83)	0.530
Direct bilirubin (μmol/L)[b]	4.95 (3.50, 10.10)	5.60 (3.90, 47.50)	< 0.001	5.10 (3.53, 9.50)	4.60 (3.50, 8.43)	0.564
Albumin (g/L)[b]	42.45 (38.30, 45.50)	43.95 (40.40, 46.40)	0.033	42.70 (38.48, 46.10)	43.30 (40.03, 46.00)	0.682
Globulin (g/L)[b]	25.40 (23.00, 27.60)	26.20 (23.80, 28.50)	0.089	25.40 (23.20, 27.83)	25.60 (23.53, 28.10)	0.533
Tumor size (cm)			0.010			0.901
≤ 3	62 (39.2)	126 (52.9)		57 (43.5)	59 (45.0)	
> 3	96 (60.8)	112 (47.1)		74 (56.5)	72 (55.0)	
CA 19-9 (U/mL)			0.378			0.537
≤ 300	75 (47.5)	125 (52.5)		64 (48.9)	70 (53.4)	
> 300	83 (52.5)	113 (47.5)		67 (51.1)	61 (46.6)	

Except where indicated, data are numbers of patients, with percentages in parentheses. NAT, neoadjuvant therapy; CA 19-9, carbohydrate antigen 19-9;

Based on the latest, most comprehensive matching of pre-operative factors, we have re-conducted the survival analysis:“ *Survival analyses stratified by the model-defined risk groups revealed that high-risk patients who received NAT followed by surgery experienced significantly longer OS compared to those treated with upfront surgery (median OS: 17.4 vs. 34.1 months; HR: 1.87, 95% CI: 1.27-2.76, P < 0.001) (Fig. 4c). This survival benefit remained significant after PSM (median OS: 18.2 vs. 34.1 months; HR: 1.91, 95% CI: 1.16-3.14, P = 0.004) (Fig. 4d). In contrast, among low-risk patients, NAT did not confer a significant OS advantage before or after matching (all P > 0.05) (Fig. 4a and 4b).* ”

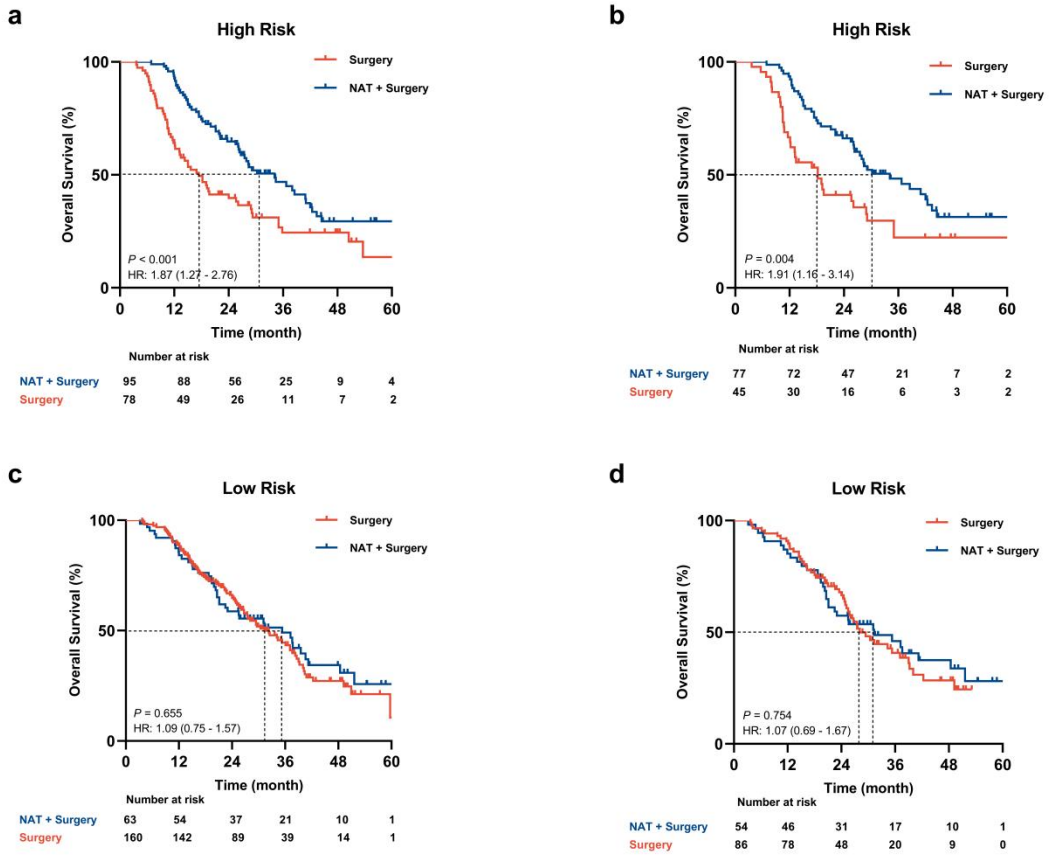
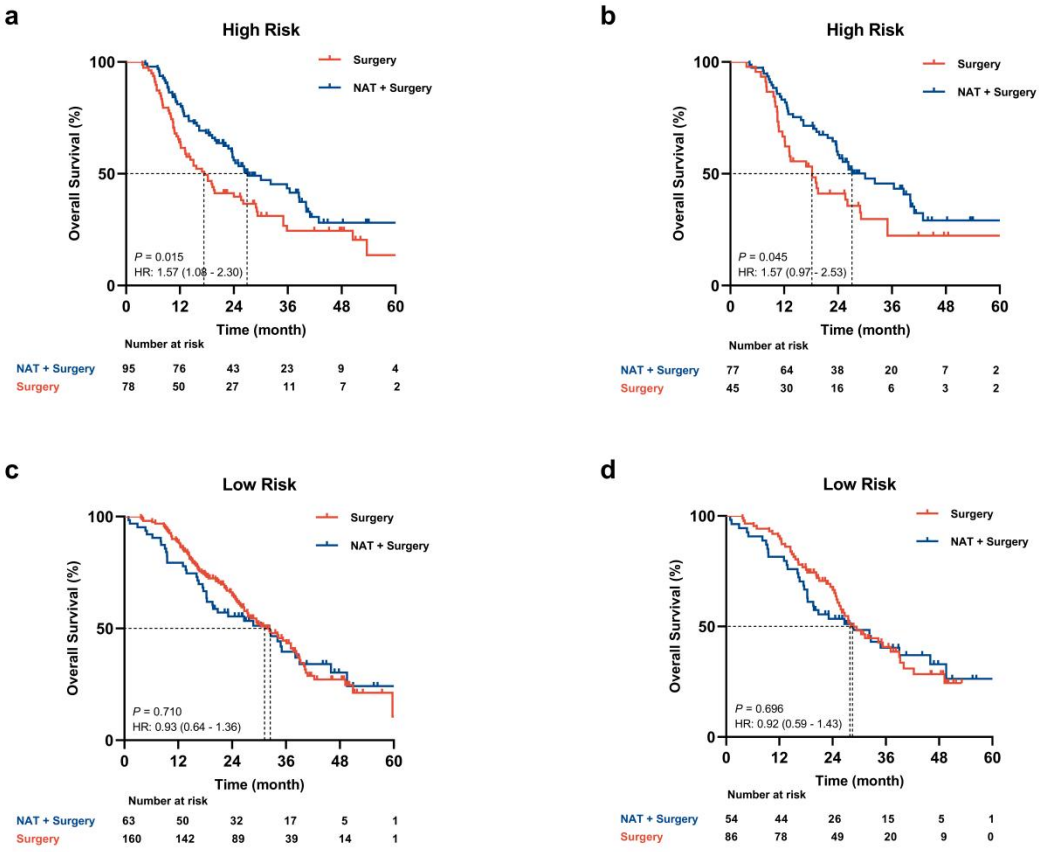


Fig. 4 | Mamba-based model predicts benefit from neoadjuvant chemotherapy. Kaplan-Meier curves of OS for patients who received or did not receive NAT in different risk groups stratified by the Mamba-based model. The findings in the total cohort (**a**, **c**) versus the results in the cohort after propensity score matching (**b**, **d**). OS, overall survival; NAT, neoadjuvant therapy

While it is difficult to include all patients who received chemotherapy without surgery, we also compared survival outcomes for patients who underwent surgery following NAT (with survival time calculated from the start of surgery) against those who had upfront surgery. The results also demonstrated consistent findings:“ *The landmark OS analysis, which calculated survival from the time of resection for the NAT group, also supported these findings, showing a consistent survival advantage for high-risk patients receiving NAT (Extended Data Fig. 8).* ”



Extended Data Fig. 8 | Sensitivity analysis of Mamba-based model predicts benefit from neoadjuvant chemotherapy. Kaplan-Meier curves of landmark OS for patients who received or did not receive NAT in different risk groups stratified by the model. The findings in the total cohort (**a**, **c**) versus the results in the cohort after propensity score matching (**b**, **d**). OS, overall survival; NAT, neoadjuvant therapy

We once again sincerely thank the reviewer for the valuable suggestions. Indeed, obtaining such data is quite challenging, and we have also incorporated as comprehensive an analysis as possible regarding the interesting findings related to the neoadjuvant therapy subgroup within our model. We hope to have the opportunity to conduct prospective cohort studies in the future to provide higher-level evidence in this area.

Comment #3:

Independent contribution of liver-derived imaging features

A central premise of the manuscript is that liver imaging features provide complementary information beyond the primary tumor. While this is supported conceptually by the proposed architecture, more systematic evidence is needed to demonstrate their independent contribution.

In particular, clearer ablation analyses comparing tumor-only, liver-only, and combined tumor–liver models would be helpful. Given the multi-center nature of the dataset, it is also important to exclude the possibility that liver features are inadvertently capturing center-specific imaging characteristics rather than biologically meaningful signals.

Reesponse：

Thank you for this insightful comment. We acknowledge the importance of providing more systematic evidence for the independent contributions of liver imaging features. To address this concern, we conducted a comprehensive ablation study comparing three distinct models.

Tumor-only model: Only the three-phase pancreatic tumor images were input into the encoder. The cross-attention fusion module was removed, and classification was performed based solely on tumor-derived features.

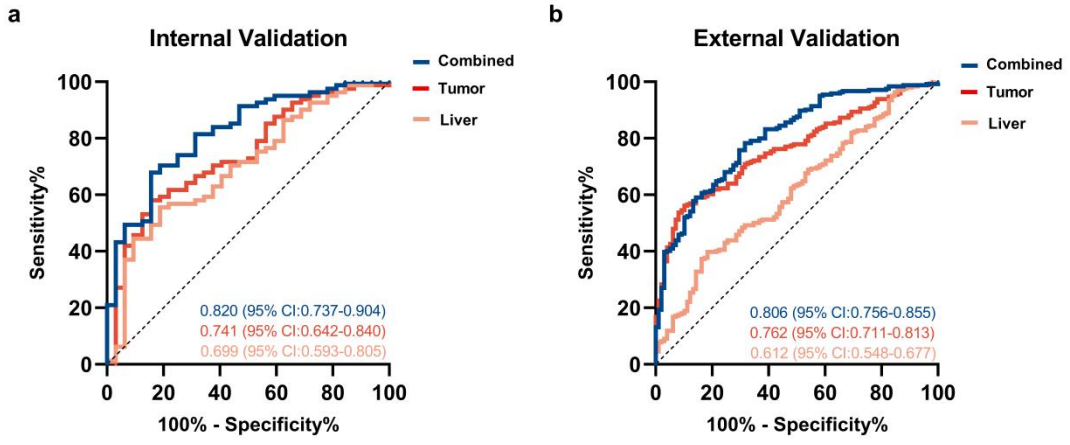
Liver-only model: Similarly, only three-phase liver images were used as input, without the

cross-attention fusion module.

Combined tumor-liver model (combined model): Both branches were jointly trained with the proposed cross-attention-based feature interaction module.

All models were trained and evaluated under identical data splits, preprocessing, and training settings to ensure fair comparison.

The results of this ablation study clearly demonstrate the complementary value of incorporating both tumor and liver features. As shown in the table below, the combined tumor-liver model consistently achieved superior performance compared to the individual component models across both the internal validation and external validation datasets. This substantial improvement confirms that the information derived from the liver parenchyma provides complementary value beyond the primary tumor features, supporting the core premise of our study. To address this, we have supplemented the manuscript with these findings:" *Albation study also demonstrated that the integrated information from both the primary tumor and the liver parenchyma was crucial for optimal predictive performance (Extended Data Fig. 3).*"



Extended Data Fig. 3 | Ablation analyses comparing tumor-only, liver-only, and combined models

Receiver operating characteristic curves in internal validation (**a**) and external validation cohort (**b**)

Additionally, to provide a more comprehensive comparison and highlight the superior performance of our proposed combined model, we have included detailed predictive performance metrics for all three models (Combined, Tumor-only, Liver-only) on the external validation cohort. Crucially, the Combined model demonstrated a significant improvement in sensitivity.

	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Combined	0.806 (0.756-0.855)	0.754 (0.709-0.800)	68.4% (63.4%-73.3%)	75.4% (73.9%-82.6%)
Tumor-only	0.762 (0.711-0.813)	0.696 (0.649-0.743)	49.0% (38.7%-59.5%)	77.9% (72.1%-82.8%)
Liver-only	0.612 (0.548-0.677)	0.556 (0.503-0.608)	56.1% (46.2%-66.3%)	55.3% (49.0%-61.4%)

Comment #4:

Generalizability of the model

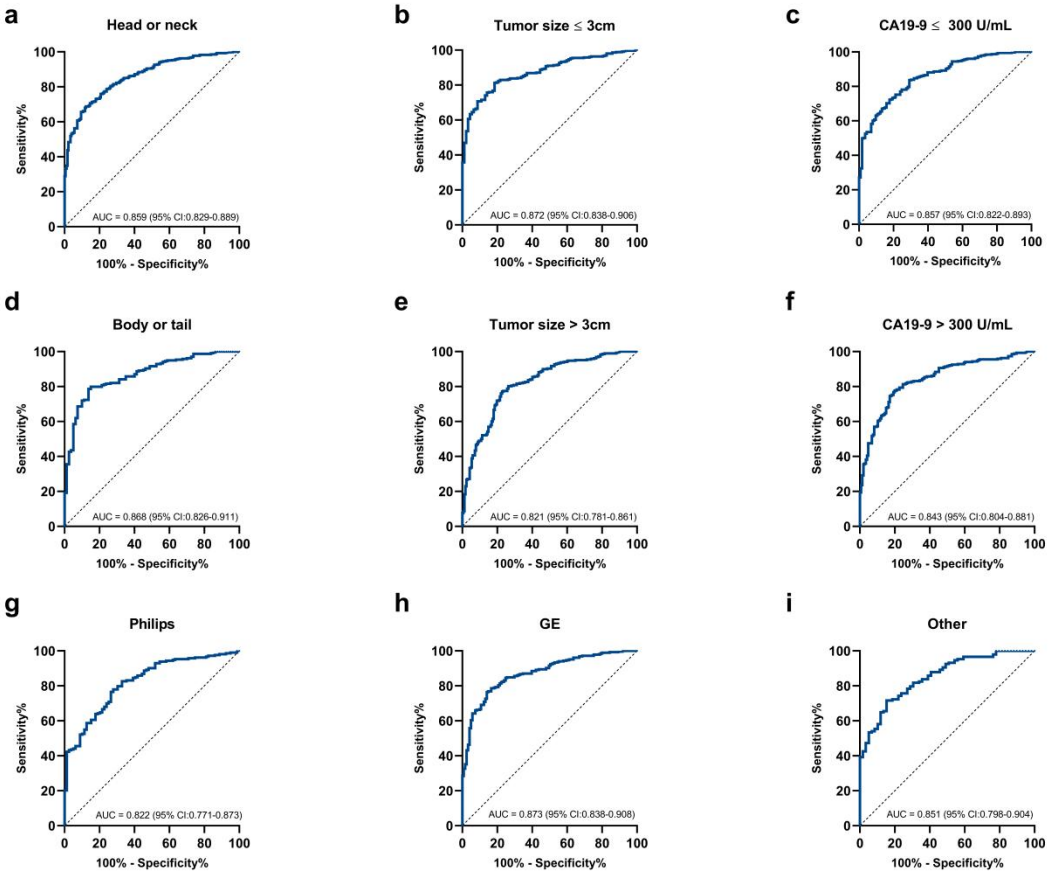
Although an external validation cohort is included, all centers are drawn from a relatively homogeneous geographic and ethnic population. This limitation should be more clearly acknowledged when discussing potential clinical deployment. Additional subgroup analyses stratified by scanner vendor or imaging parameters would further support claims of robustness.

Reesponse：

We thank the reviewer for this valuable comment regarding the generalizability of our model. We acknowledge that the patient population in our study was derived exclusively from East Asian centers. We have addressed this limitation explicitly in the Limitations of the revised manuscript:“ *Second, the patient population was derived exclusively from East Asian centers and was relatively homogeneous in terms of geography and ethnicity. The relatively high incidence of ELM in our study may limit the generalizability of our findings to other ethnic or geographic populations with different baseline clinical characteristics.”*

To comprehensively evaluate the model’s performance across diverse clinical settings, we

expanded our subgroup analyses beyond the initial clinical scenario stratifications. Specifically, we performed additional subgroup analyses stratified by scanner vendor to assess the model's robustness under varying technical conditions:" *To further assess generalizability, we conducted subgroup analyses stratified by tumor location, tumor size, and preoperative CA19-9 levels. The model maintained consistently high discriminatory ability across subgroups, with AUCs ranging from 0.821 to 0.872, suggesting its broad clinical applicability. Furthermore, the model demonstrated stable performance with AUCs remaining consistently high with different scanner vendors (Extended Data Fig. 4), suggesting its robustness across varied technical settings.* "



Extended Data Fig. 4 | The performance of Mamba-based model to predict the early liver metastases in different subgroups. Receiver operating characteristic curves in different subgroups defined by tumor location (**a** and **d**), tumor size (**b** and **e**), CA 19-9 levels (**c** and **f**) and different

scanner vendors (**g-i**)

Comment #5:

More detail on the quality control of automated segmentations would improve transparency, including any assessment of inter-reader variability during manual correction.

Response:

Thank you for this important suggestion. We acknowledge that providing more details on the quality control of automated segmentations enhances the transparency and reproducibility of our methodology. As requested, we have included additional information regarding the assessment of inter-reader variability during the manual refinement process.

Specifically, to evaluate the consistency and accuracy of the manual corrections applied to the automated segmentations, we randomly selected 100 patients. Two radiologists independently reviewed and refined the segmentations for the two key volumes of interest: the pancreatic cancer lesion and the whole liver. The Dice Similarity Coefficient (DSC) was calculated to quantify the overlap between the segmentations produced by the two radiologists. The resulting average DSCs were 0.94 for the pancreatic cancer lesions and 0.99 for the liver segmentations, indicating excellent inter-reader agreement and high-quality manual refinements guided by the automated model.

We have incorporated this information into the Methods section under the subsection detailing the segmentation procedure:“ *These masks were subsequently refined manually by radiologists to ensure high accuracy. The inter-rater reliability of these manual refinements yielding Dice Similarity Coefficients (DSCs) of 0.94 for pancreatic lesions and 0.99 for liver segmentations, indicating excellent agreement between readers and validating the quality of the*

refined masks."

Comment #6:

The risk threshold is defined using the Youden index; reporting additional clinically interpretable operating points (e.g., fixed sensitivity or negative predictive value) would enhance translational relevance.

Response:

Thank you for this insightful comment regarding the clinical interpretability of our model's risk threshold. Given the clinical priority often placed on high sensitivity to minimize false negatives, we identified an operating point targeting approximately 90% sensitivity. An operating point closest to 90% sensitivity was identified at a threshold of 0.266, yielding a sensitivity of 89.9% (95% CI: 83.4-94.5%) and a specificity of 72.6% (95% CI: 67.4-77.4%). The corresponding positive likelihood ratio (+LR) was 3.28, and the negative likelihood ratio (-LR) was 0.14.

We primarily reported the threshold derived from the Youden index in the main text, as it represents the optimal balance between sensitivity and specificity. Notably, the Youden index-derived threshold already achieved a high sensitivity of 87.6% (approaching our target of 90%), while maintaining a higher specificity (78.8%) compared to the 90%-sensitivity threshold (72.6%). Therefore, we reported only the threshold derived from the Youden index.

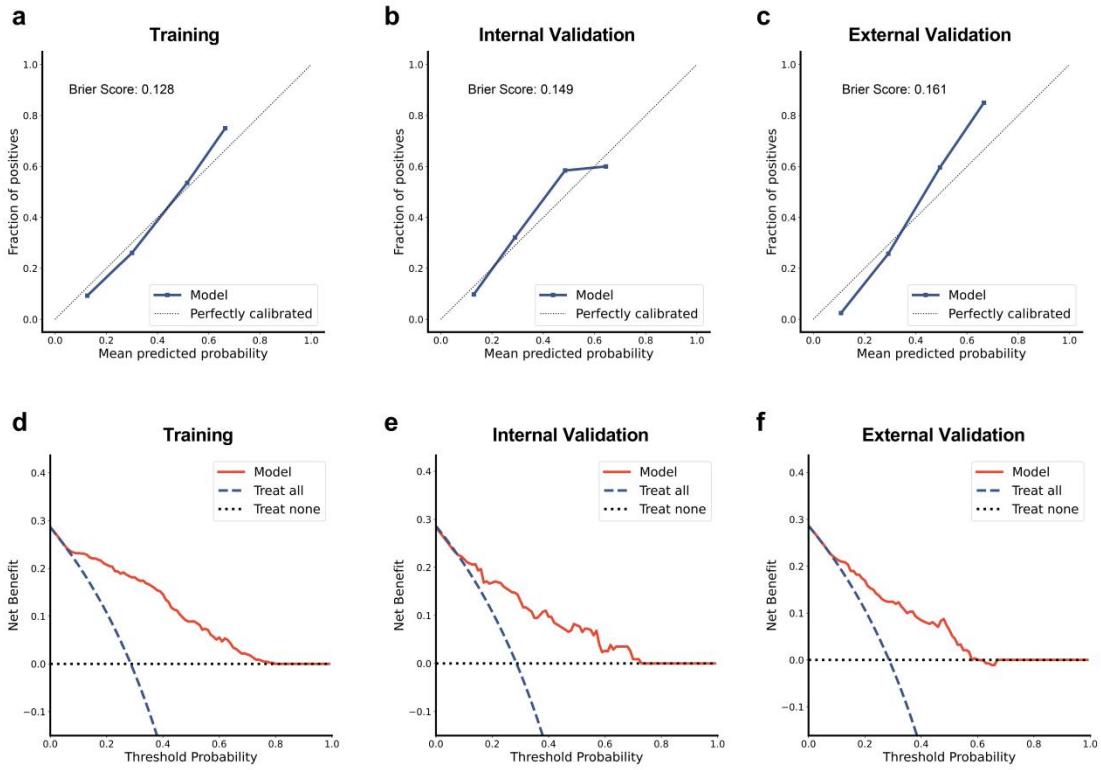
Comment #7:

Calibration performance is mentioned but could be more fully quantified, particularly in the external validation cohort.

Response:

Thank you for this important observation regarding the calibration performance. To address this

concern, we have substantially strengthened the calibration analysis. Firstly, we have updated the calibration curves to include the Brier Score (Extended Data Fig. 5). Furthermore, to provide a more rigorous statistical assessment of calibration across validation cohorts, we have included the following results in the main text: " *Moreover, calibration curves showed excellent agreement between predicted probabilities and actual outcomes across all cohorts, indicating strong model reliability. The Hosmer-Lemeshow goodness-of-fit test yielded non-significant P-values in the validation cohorts (P ≥ 0.116), confirming adequate calibration.*"



Extended Data Fig. 5 | Calibration curves and decision curve analysis. The calibration curves for the Mamba-based model in the training and validation cohorts (a-c); The decision curve analysis for the Mamba-based model in the training and validation cohorts (d-f)

Comment #8:

The radiotranscriptomic analysis is based on a relatively small sample size and should be framed

more explicitly as exploratory.

Response:

We thank the reviewer for this important observation. We acknowledge that the radiotranscriptomic analysis presented in our study involved a relatively small sample size. As suggested by the reviewer, we agree that this component of our work should indeed be considered and explicitly framed as exploratory. We revised our manuscript as follows:" *To elucidate the underlying biological basis of the Mamba-based model in predicting ELM, we performed an exploratory radiotranscriptomics analysis on a subset of patients from the ××× (Center 5) with matched preoperative CECT and bulk RNA sequencing data. For each patient, a risk score was generated using the trained model, and patients were stratified into high- and low-risk groups accordingly."*

Comment #9:

Additional details on preprocessing and inference pipelines would further improve reproducibility.

Response:

We thank the reviewer for emphasizing the importance of providing additional details on the preprocessing and inference pipelines to enhance reproducibility. We have now clarified these procedures in the revised manuscript.

Preprocessing step：

Briefly, given a 3D abdominal CT volume, preprocessing was performed following the self-configuring pipeline of nnU-Net (v2). Each volume was first read using NibabelIO and reoriented to a consistent axis convention to ensure spatial alignment across cases. A 3D cascade resampling framework was then adopted to address anisotropic voxel spacing. For the

full-resolution stage, volumes were resampled to 2.30 × 0.72 × 0.72 mm^3 for pancreatic tumor analysis and 2.50 × 0.73 × 0.73 mm^3 for liver analysis. In the low-resolution stage, the spacing was set to 2.91 × 1.46 × 1.46 mm^3 for pancreatic tumor and 2.50 × 1.28 × 1.28 mm^3 for liver. Third-order spline interpolation (order = 3) was applied to image intensities, while linear interpolation was used for segmentation masks, with the z-axis handled separately due to anisotropy. After resampling, each volume was cropped to the bounding box of non-zero voxels to remove irrelevant background regions and reduce memory consumption.

Intensity normalization was subsequently performed based on foreground statistics computed from the training set. Voxel intensities were clipped to the 0.5th and 99.5th percentiles of foreground voxels, corresponding to [1, 135] HU for pancreatic tumor and [13, 117] HU for liver in our dataset. Z-score normalization was then applied using the foreground mean and standard deviation (pancreatic tumor: mean 65.12, std 25.51; liver: mean 66.21, std 17.99), which helps mitigate inter-center intensity variability.

In the second stage, a pretrained nnU-Net segmentation model was used to obtain tumor and liver masks. These masks were subsequently refined manually by radiologists to ensure high accuracy. Corrected masks were then multiplied with the preprocessed CT volumes to isolate organ-specific regions. Center-based cropping was then performed to generate fixed-size inputs for classification, with tumor regions cropped to 128 × 128 × 32 and liver regions to 256 × 256 × 64. The three-phase tumor and liver volumes were saved in .npy format for downstream model input.

Inference pipeline：

During inference, each case undergoes the identical preprocessing pipeline described above

to ensure consistency with training. The resulting three-phase tumor and liver .npy files are organized according to the directory structure specified in our GitHub repository and fed into the trained dual-branch network to produce binary classification probabilities. For reproducibility, we provide representative sample data in the data/ directory, including preprocessed .npy files of the non-contrast, arterial, and venous phases for both tumor and liver regions. The data organization during inference follows the same structure as in the training stage, with the three-phase images placed separately in the noncontrast/, arterial/, and venous/ subdirectories. We also release the inference scripts (inference.py and inference_nolabel.py) in the repository to facilitate direct evaluation and deployment. The trained model weights have been already uploaded in the Google Drive (https://drive.google.com/drive/folders/1J9L-5HTu3gXzFG5o3Bs-VoF66-6rph0h). The download link and inference pipeline is also updated in the README file of the repository to enhance reproducibility.

We have added a description of the preprocessing steps in the main text:“ *Given the superior sensitivity of the arterial phase for detecting PDAC and delineating tumor margins, non-contrast and portal venous phase images were co-registered to the arterial phase to enhance spatial consistency. Subsequently, a standardized preprocessing pipeline adapted from the self-configuring nnU-Net (v2) framework was applied. This included reorientation, resampling, and intensity normalization based on foreground statistics derived from the training set. Following preprocessing, automated segmentation of the primary tumor and liver regions were performed using the nnU-Net framework with default parameters 45. These masks were subsequently refined manually by radiologists to ensure high accuracy. The inter-rater reliability of these manual refinements was assessed on a subset of cases, yielding Dice Similarity Coefficients (DSCs) of*

0.94 for pancreatic lesions and 0.99 for liver segmentations, indicating excellent agreement between readers and validating the quality of the refined masks. ” also revised the inference details:“ *During the preprocessing stage, a 3D cascaded resampling framework was adopted to address anisotropic voxel spacing. In the full-resolution stage, volumes were resampled to 2.30 × 0.72 × 0.72 mm³ for pancreatic tumor analysis and 2.50 × 0.73 × 0.73 mm³ for liver analysis. In the low-resolution stage, the spacing was set to 2.91 × 1.46 × 1.46 mm³ for pancreatic tumors and 2.50 × 1.28 × 1.28 mm³ for liver. Voxel intensities were clipped to the 0.5th and 99.5th percentiles of foreground voxels, corresponding to [1, 135] HU for pancreatic tumors and [13, 117] HU for liver in our dataset. Subsequently, Z-score normalization was applied using the foreground mean and standard deviation (pancreatic tumor: mean = 65.12, std = 25.51; liver: mean = 66.21, std = 17.99). After segmentation, center-based cropping was performed to generate fixed-size inputs for the classification network. Tumor regions were cropped to 128 × 128 × 32, while liver regions were cropped to 256 × 256 × 64.*

During training, several data augmentation strategies were applied to improve model robustness, including random noise injection, random bias field distortion, random flipping, and random motion. The model was optimized using the Adam optimizer with a binary cross-entropy loss function. Training was conducted for 100 epochs with an initial learning rate of 1e−5, which was gradually decreased to 1e − 7 using a cosine annealing schedule. All experiments were performed on a single NVIDIA RTX 4090 GPU.” Comprehensive implementation details, including code usage and environment setup, are provided in the accompanying README file.

Comment #10:

The authors provide a publicly accessible GitHub repository with source code and a README

describing the overall pipeline. The repository is generally well organized and facilitates understanding of the proposed architecture.

However, full reproducibility is currently limited by the absence of example data, pretrained weights, and detailed instructions for reproducing the reported experiments across cohorts. Clarifying the preprocessing steps, model configuration files, and inference workflow would further enhance the utility of the code as a community resource.

Response:

Thank you for the positive feedback and recognition regarding our code availability. We appreciate your suggestion to enhance reproducibility.

The detailed preprocessing and inference procedures have been described in our response to Comment #9, and the specific details have been incorporated into the README file accompanying the code repository, which describes the overall pipeline. Furthermore, as suggested, we have provided the model configuration files.

We believe these updates address your concerns and enhance the utility of the code as a community resource. If there are any further questions or areas requiring clarification, please do not hesitate to contact us. We are grateful for your valuable suggestions, which have significantly improved the quality of our manuscript.

First-round review comments

Response to Reviewer #3:

General Comments:

This study evaluated the utility of Mamba-based predictive model integrating imaging features from pancreatic tumor and the liver to determine the risk of occult liver metastases in a multi-institutional cohort. The model showed good performance in risk stratification and in predicting patient survival. In a subgroup of patients, the radiotranscriptomic analysis provided plausible biologic explanation of the model.

Overall Response:

We sincerely thank the reviewer for the positive evaluation and high recommendation of our study. We greatly appreciate the thoughtful and detailed feedback regarding the definition of occult liver metastasis and various aspects of manuscript presentation. In response, we have carefully addressed all comments in a point-by-point manner. The manuscript has been substantially improved through this rigorous review process, and we are grateful for the reviewer's insightful suggestions and meticulous reading of our work.

Comment #1:

The definition of liver metastases within 6 months postop as "occult liver metastasis" is not a commonly accepted definition. Occult liver metastasis usually refers to detection of liver metastasis during surgical exploration that occur soon (within 1 month) of the CT. 6 months is a long window of opportunity for the liver to develop metastases. Consider using other terms such as "early liver recurrence" or "high risk for liver metastases" to convey the idea that the patients are at high risk, but the liver metastases were not necessarily present on the CT.

Response:

Thank you for your insightful comment. We agree that the term “occult liver metastasis” lacks a universally standardized definition. To improve clarity and avoid misinterpretation, we have revised the terminology throughout the manuscript. As suggested, we now use the term “early liver recurrence” to describe metastases detected within 6 months postoperatively. This better reflects the clinical scenario.

Comment #2:

The rate of “occult liver metastases” ~28% (within 6 months) seem high, based on published literature and the experience at my institution. Median recurrence for upfront surgery is ~12-18 months. Therefore, this model may not be applicable at other sites.

Response:

Thank you for your positive comments and questions. Indeed, the definition of early liver metastases and reported rates differ in literatures, ranging from approximately 20% within 6 months in some studies to as high as 45% within 3 months in others[1,2]. These discrepancies across institutions are likely attributable to differences in imaging protocols, surveillance strategies, and clinical practices.

The rate of “early liver metastases” ~28%, which seems slightly higher. We provide the following potential explanations for the finding: many patients with PDAC, especially those with ELM, often present with early symptoms or biochemical abnormalities, they tend to undergo postoperative imaging studies early and repeatedly. This practice may increase the observed proportion of metastatic disease in our data. Conversely, individuals who are asymptomatic or lost to systematic follow-up may be under-represented because timely imaging and regular surveillance are lacking, and partial of these patients without regular follow-up may be excluded.

Consequently, our reported rate of liver metastasis may be probably over-estimated. We also explicitly acknowledge this selection bias as a limitation of the present study: *"Second, the patient population was derived exclusively from East Asian centers, and the relatively high incidence of ELM in our population may limit the generalizability of findings to other ethnic or geographic populations with different clinical characteristics."*

Reference

1. Wang, F. *et al.* Evaluation of occult liver metastases in pancreatic adenocarcinoma by diffusion-weighted related magnetic resonance imaging. *J Magn Reson Imaging* **62**, 536–548 (2025).

2. Liu, J.-B. *et al.* Development and validation of a predictive model for occult liver metastasis in pancreatic ductal adenocarcinoma using subjective imaging and clinical data. *Cancer Med* **14**, e71280 (2025).

Comment #3:

The patients who underwent upfront surgery vs. neoadjuvant therapy where not randomized, which may introduce a source of bias, despite the propensity analysis.

Response:

We sincerely thank the reviewer for this insightful comment. Although our dataset represents one of the largest multi-institutional cohorts with detailed imaging, treatment, and outcome data. We fully acknowledge that the non-randomized nature of treatment allocation in this retrospective study may introduce potential selection bias, even after rigorous propensity score matching. As suggested, we have explicitly addressed this limitation in the revised manuscript: "*First, despite the use of propensity score matching and multivariable adjustments, the retrospective study designinherently introduces potential selection bias and unmeasured confounding,which may influence*

the observed associations." Looking forward, we agree that higher-level evidence is needed to confirm the clinical utility of our model in guiding treatment selection. As such, we have also added a statement: "*Finally, our findings require prospective validation in large and multicenter cohorts. Future studies should further explore the utility of this model in guiding other treatment strategies, such as adjuvant therapy intensity or immunotherapy selection, and assess its real-time clinical integration in multidisciplinary treatment workflows.*"

Comment #4:

Introduction. Statement of "Over half of these hepatic recurrences occur within 6 months" from References 5, 9. These references did not seem to state that. Ref 5 cites a Cochrane review describing staging laparoscopy detecting unresectable disease following CT (the usual definition for occult metastatic disease). Please clarify.

Response:

We sincerely appreciate your attention to detail and the opportunity to improve the clarity and accuracy of our manuscript. As you mentioned, these 2 reviews state PDAC is an aggressive disease, typically with early systemic spread, and most patients eventually develop systemic disease recurrence, suggesting the presence of systemic micrometastases. These references do not specifically support the claim that "over half of hepatic recurrences occur within six months postoperatively." We acknowledge that the original citation was therefore inappropriate, and we apologize for this inaccuracy.

To better reflect the current understanding and to enhance the precision of our statement, we have revised the text as follows: "*Yet, disease recurrence is distressingly common, affecting up to 80% of patients. The liver is the most frequent site of distant recurrence, with a substantial*

proportion of liver recurrences occurring in the early postoperative period, often within six months after curative-intent resection [6–8]. These early liver metastases are often characterized by aggressive biological behavior, likely reflecting the presence of occult micrometastatic disease, and are associated with dismal post-recurrence survival, typically less than six months [6,9,10]." We hope the revised statement can more accurately capture the clinical reality of early liver metastasis in PDAC, align with the cited references, and improve the contextual coherence of our discussion.

Reference

6. Zhang, X.-P. *et al.* Early and late recurrence patterns of pancreatic ductal adenocarcinoma after pancreaticoduodenectomy: a multicenter study. *Int J Surg* **109**, 785–793 (2023).

7. Groot, V. P. *et al.* Defining and Predicting Early Recurrence in 957 Patients With Resected Pancreatic Ductal Adenocarcinoma. *Ann. Surg.* **269**, 1154–1162 (2019).

8. Groot, V. P. *et al.* Patterns, timing, and predictors of recurrence following pancreatectomy for pancreatic ductal adenocarcinoma. *Ann Surg* **267**, 936–945 (2018).

9. Groot, V. P. *et al.* Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. *Annals of Surgery* **267**, 936–945 (2018).

10. Tsilimigras, D. I. *et al.* Liver metastases. *Nat Rev Dis Primers* **7**, 27 (2021).

Comment #5:

2. Introduction. Neoadjuvant therapy is meant to target patients with high-risk disease, not just patients with subclinical liver disease.

Response:

You are absolutely correct. Neoadjuvant therapy (NAT) is increasingly used to treat patients with high-risk features beyond just subclinical liver involvement—such as borderline resectability,

elevated CA19-9, large primary tumors, or suspicious lymphadenopathy.We have revised the Introduction to reflect this broader rationale: "*Neoadjuvant therapy (NAT) represents a promising therapeutic approach for high-risk PDAC patients, particularly those likely with high risk of occult micrometastatic disease.*"

Comment #6:

2. Results. Please add preoperative local staging (resectable, borderline resection, unresectable) and progression free survival and overall survival for the whole cohort and subgroups in the text or tables. The local staging is especially relevant for the patients from Center 5 who underwent upfront surgery or NAT.

Response:

We sincerely appreciate the reviewer's insightful comment. As suggested, we have supplemented the preoperative local staging (resectable, borderline resectable), as well as progression-free survival (PFS) and overall survival (OS) for the entire cohort and relevant subgroups, which are now clearly presented in Table 1.

While all enrolled patients were classified as having localized disease (including both resectable and borderline resectable categories) at baseline based on multidisciplinary assessment, we fully agree with the reviewer that local staging is a critical determinant of clinical decision-making, particularly for patients from Center 5, who received either upfront surgery or neoadjuvant therapy (NAT). To address this, we have incorporated local staging as a covariate in Extended Data Table 4.

Besides, to minimize selection bias and confounding due to these imbalances in baseline prognostic factors, we performed propensity score matching (PSM) using key clinical variables,

including age, sex, tumor size, CA19-9 level, and local staging. After matching, 123 patients remained in both the surgery-first and NAT groups. In this balanced cohort, we observed that model-defined high-risk patients who received NAT followed by surgery experienced significantly longer OS compared to those treated with upfront surgery (median OS: 17.4 vs. 34.1 months; HR: 1.87, 95% CI: 1.27–2.76, $P < 0.001$) (Fig. 4c). This survival benefit remained significant after PSM (median OS: 17.8 vs. 30.2 months; HR: 1.71, 95% CI: 1.04–2.80, $P = 0.019$) (Fig. 4d).

Comment #7:

4. Results P.9 line 1. Typo in “before or after mating”. Should be “matching”.

Response:

Thank you for identifying this typographical error. We apologize for the oversight. The text has been corrected from “before or after mating” to “*before or after matching*” in the revised manuscript.

Comment #8:

Methods. Liver segmentation. Did the segmentation exclude benign lesions (e.g., cysts, hemangiomas) that may affect the imaging features of the liver?

Response:

Thank you for your insightful comment regarding the impact of benign liver lesions on model’s performance. In our model, the liver was included as a key input via a dual-branch architecture, with whole-liver segmentation performed without exclusion of benign lesions (e.g., cysts, hemangiomas).

Our Mamba-based model is specifically designed to integrate spatiotemporal information from multiphase CT scans, enabling it to capture dynamic enhancement patterns across both the

primary tumor and the liver. While benign hepatic lesions, predominantly cyst, hemangiomas and calcifications, do exhibit distinct enhancement kinetics, our framework effectively learns to differentiate these from malignancy-associated patterns through end-to-end training. Besides, we regard the presence of benign lesions as a natural perturbation that enhances model robustness in real-world settings.

To further address your concern, we conducted a supplementary analysis in a subset of patients with benign lesions manually excluded. This refinement did not yield a meaningful improvement in diagnostic performance, with AUCs of 0.882 (training), 0.803 (internal validation), and 0.780 (external validation), while substantially increasing segmentation burden. Given the minimal gain in accuracy and the added clinical impracticality, we recommend whole-liver segmentation as a robust and efficient strategy for model input. We appreciate this valuable comment, which prompted us to further validate our methodological choice.

Comment #9:

Table 1, extended Table 1-3: “Resection region” should be changed to “Resection margin”.

Response:

Thank you for your careful review and this accurate observation. We agree that “Resection region” was incorrect. The term has been revised to “*Resection margin*” in Table 1 and Extended Data Tables 1–3 to accurately reflect the pathological assessment.

Comment #10:

The proportion of patients who did not undergo adjuvant treatment seemed high in the internal cohorts (23.8-25.7%), and that may have an important impact on survival.

Response:

Thank you for your positive comments and questions. Thank you for this important observation. Indeed, the proportion of patients not receiving adjuvant therapy in the internal cohorts (23.8 – 25.7%) reflects real-world clinical heterogeneity, which arises from differences in postoperative complications, performance status, patient preference, and access to care.

PDAC is a highly aggressive and clinically heterogeneous disease, and treatment patterns naturally vary across institutions. Our multicenter design captures this diversity, and we did not enforce treatment homogeneity across sites to preserve the real-world generalizability of our findings. Importantly, Survival analyses were performed within each cohort by first stratifying patients into high- and low-risk groups according to our model's output, and then comparing survival outcomes between these strata and we consistently observed significant survival differences favoring the low-risk group across all cohorts.

Response to Reviewer #4:

General Comments:

In this manuscript, Zhao et al. present a carefully designed study that applies the emerging Mamba architecture to predict occult liver metastasis (OLM) in pancreatic ductal adenocarcinoma (PDAC) using multi-phase CT. The proposed dual-region model, jointly analyzing the primary tumor and liver parenchyma, is biologically well-grounded as it reflects the metastatic cascade. The robust performance across multiple cohorts, together with the potential to guide neoadjuvant therapy (NAT) decisions, underscores the strong translational value of this work. This is an interesting and timely study that addresses a clinically important question with a novel AI architecture, biologically rational modeling, and strong validation. With minor clarifications and

revisions as outlined below, the manuscript will be further strengthened. Overall, this paper is suitable for publication in Nature Communications pending minor revisions.

Overall Response:

Thank you very much for your thoughtful review and constructive feedback. We greatly appreciate your careful reading of our manuscript, your recognition of its scientific merit, and your valuable suggestions for improvement. Your insightful comments have helped us strengthen the quality and clarity of the work, and we have addressed all points in detail in the revised manuscript.

Comment #1:

Choice of Mamba architecture: Please expand on the rationale for adopting the Mamba framework rather than conventional transformer- or CNN-based models. A clear explanation will highlight the novelty and technical advantages of this choice.

Response:

Thank you for your insightful comment regarding the superiority of using Mamba. Mamba was selected for its ability to explicitly model spatial and temporal contrast variations through its selective state-space modeling (SSM) mechanism. Unlike CNNs that are limited by local receptive fields or Transformers that incur quadratic complexity, Mamba employs a linear recurrent formulation to efficiently capture long-range voxel-level dependencies while maintaining scalability to high-resolution 3D data.

This design choice aligns closely with the clinical interpretive process, where radiologists evaluate temporal enhancement patterns across the liver and tumor over multiple phases. By recursively updating hidden states, Mamba achieves fine-grained spatiotemporal sensitivity and

global contextual awareness that conventional architectures struggle to balance.

Furthermore, Mamba operates with linear computational complexity, a crucial advantage given our task involves handling large-scale data input--three high-resolution 3D CT volumes across a dual-branch architecture. By efficiently managing this substantial computational burden, Mamba proves exceptionally suited to our task, enabling our method to achieve significantly lower training and inference costs compared to Transformer-based models while maintaining modeling capacity.

Comment #2:

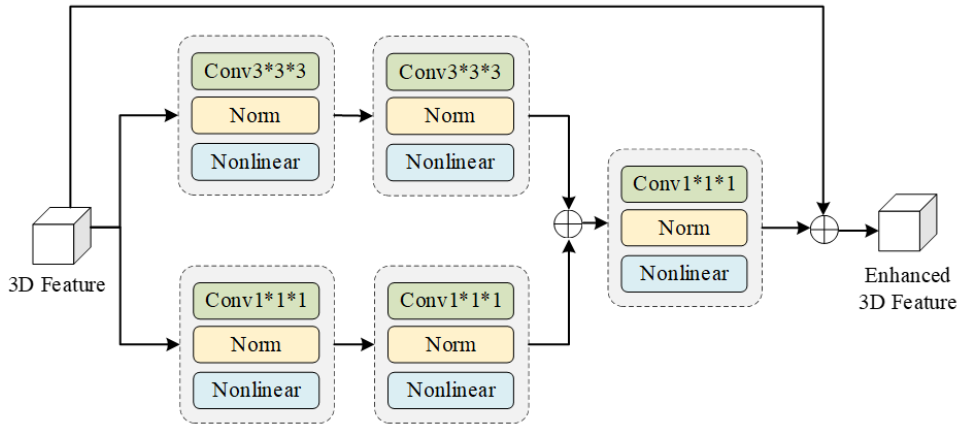
Spatial feature representation: Flattening 3D features into 1D sequences may risk losing spatial context. Could the authors clarify how Mamba's modeling capability effectively preserves and leverages the spatial information in CECT images?

Response:

Thank you for raising this important concern. We fully agree that flattening 3D features into a 1D sequence may risk losing spatial coherence. In our framework, this operation is carefully designed to enable fine-grained global modeling through dedicated spatial-scanning and temporal-scanning, while explicitly preserving spatial relationships through dedicated mechanisms.

Specifically, the flattening operation reorders multi-phase 3D features into a unified sequence with dimension (Batch Size, Channel, Slice × Width × Height), decoupling spatial and temporal contexts for separate but complementary modeling. This enables our Mamba-based model's spatial-scanning and temporal-scanning modules to capture fine-grained, phase-specific contrast-enhanced patterns across the liver and tumor region.

To preserve spatial context, the sequence is subsequently reshaped back into its original 3D lattice for further processing, ensuring structural consistency. Furthermore, to mitigate the loss of spatial information after flattening, we design a spatial-enhanced module (Supplement Figure 1) before each flattening operation that enriches the complementary details across different feature levels. Specifically, 3D features are processed through two parallel paths: One path employs two 3×3×3 convolution blocks to encode channel-wise spatial context. Each block consists of a convolution, a normalization layer, and a nonlinear activation layer. Another path applies two 1×1×1 convolution blocks to aggregate pixel-wise cross-channel context. Finally, the enhanced spatial features are obtained through residual connections. More details can be seen in the source code we provide.



Supplement Figure 1. Spatial-enhanced module in our framework that enriches the spatial information.

Comment #3:

Biological interpretability of dual branches: The two-branch design (tumor vs. liver) is conceptually elegant and biologically rational. It would strengthen the paper to discuss whether the model identifies interpretable and biologically meaningful patterns—for example, which types of features from the tumor and parenchyma branches may drive predictive performance.

Response:

We thank the reviewer for this thoughtful comment. We fully agree that biological interpretability is essential for clinical applicability. Our dual-branch design was motivated by the clinical rationale that tumor-related features (e.g., morphology, heterogeneity) and liver parenchyma features (e.g., texture, background liver condition) provide complementary biological cues for prediction. Therefore, we designed an attention-based tumor-liver fusion mechanism that adaptively balances the contribution of each branch through attention weighting. We summarize the fusion mechanism below and illustrate the challenge of direct visualization under this mechanism.

As shown in Supplement Figure 2, after the encoder stage, the three-phase features of each branch are concatenated to generate a pancreas tumor feature and a liver feature. Subsequently, each feature is refined through an intra-organ self-attention module to filter out noise and redundancy introduced by multiple downsamplings:

$$F' = \mathrm{SelfAttn}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where F' denotes the enhanced tumor or liver feature, and Q, K, V represent the query, key, and value matrices generated from the tumor/liver features. The refined features are then processed with residual connections and layer normalization:

$$F'' = \mathrm{LN}(F' + F)$$

where LN denotes the layer normalization function, F″ represents the normalized tumor feature (F_t) or liver feature (F_o), and F denotes the original feature. Next, cross-attention is applied to enable interaction between the tumor and liver features. The feature of tumor branch generates matrices of Q_t',K_t',V_t', while the liver branch generates matrices of Q_o',K_o',V_o'. The

cross-attention mechanism is defined as:

$$F_t^{int} = \mathrm{CrossAttn}(Q'_t\ K'_o\ V'_o),\ F_o^{int} = \mathrm{CrossAttn}(Q'_o\ K'_t\ V'_t),$$

where F_t^{int} and F_o^{int} denote the interaction-enhanced tumor and liver features, respectively. These features are further refined by residual connections and layer normalization:

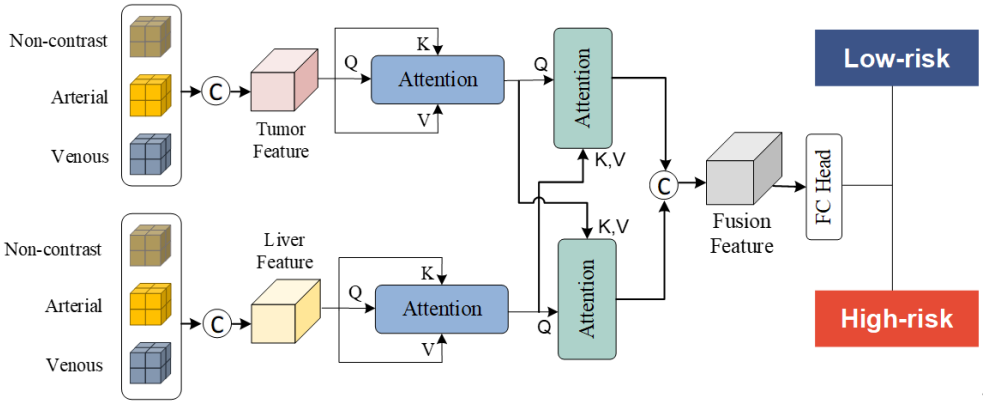
$$F''_t = \mathrm{LN}(F_t^{int} + F_t),\ F''_o = \mathrm{LN}(F_o^{int} + F_o)$$

Finally, the enhanced tumor and organ features are concatenated along the channel dimension to form the fused representation for final prediction. This design allows the model to leverage biologically complementary cues across the tumor and liver.

Experimental results demonstrate the effectiveness of this fusion mechanism. However, the designed fusion architecture poses inherent challenges for direct feature visualization (e.g., via Grad-CAM). First, both tumor and liver features are constructed by concatenating three-phase (non-contrast, arterial, and portal venous) representations. In subsequent processing, these three phases are modeled jointly as an integrated feature through intra-organ and cross-organ attention. In contrast, visualization requires phase-specific feature correspondence with the respective CT phases, which our fusion scheme does not explicitly preserve. Secondly, both the tumor and liver features are flattened into sequences and successively processed by intra-organ self-attention and cross-attention modules. These operations disrupt the original voxel-wise spatial correspondence, making it difficult to project attention weights back to the 3D input domain. Furthermore, the incorporation of residual connections and layer normalization introduces both linear and nonlinear feature mixing, diffusing the contribution of individual neurons across the entire sequence. At the final decision stage, the tumor and liver representations are concatenated along the channel dimension, forming a fused high-dimensional embedding that cannot be directly disentangled into

organ-specific spatial maps.

To address this challenge and enhance the biological plausibility of our model, we proactively incorporated complementary multi-omic analyses during the study design. Specifically, we performed radiotranscriptomic analysis by integrating imaging phenotypes with transcriptomic profiles from matched tumor tissues to reveal biologically meaningful associations.



Supplement Figure 2. The fusion mechanism of the pancreatic tumor branch and liver branch.

Comment #4:

Clinical heterogeneity across cohorts: Since NCCN guidelines recommend NAT for high-risk features such as borderline resectable disease or elevated CA19-9, it would be helpful to address potential clinicopathological heterogeneity between the development and application cohorts.

Response:

Thank you for this insightful comment. We agree that clinical heterogeneity between the development cohort and the NAT+surgery application cohort. To address this concern, we have presented a comprehensive comparison in Extended Data Table 3, which is described in the main text under the section "Assessment of the model's utility in guiding neoadjuvant therapy." This table highlights differences in key preoperative variables, including resectability status and CA19-9 levels, reflecting the NAT+surgery application cohort with higher-risk.

Also, to minimize potential bias arising from these imbalances, we subsequently performed propensity score matching (PSM) to balance the distribution of important prognostic factors between groups. We believe these additions strengthen the validity of our findings and provide a more nuanced understanding of the model's applicability in clinically heterogeneous populations.

Comment #5:

Multi-modality integration: The integration of transcriptomic and radiomic findings is a commendable strength of this work. Could the authors comment on why a more formal multi-modality integration approach was not undertaken?

Response:

We appreciate the reviewer's insightful comment regarding multi-modality integration. While formal fusion of multi-omics data is promising, our study focuses on a clinically urgent question: identifying patients with early liver metastases who may not benefit from upfront surgery. In clinical practice, preoperative imaging is the most accessible, rapid, and cost-effective tool for risk stratification, whereas transcriptomic and pathological data are typically available postoperatively. Therefore, our primary goal was to develop a preoperative predictive model using widely available radiomic features to guide clinical decision-making.

Nonetheless, we performed radiotranscriptomic analysis to provide biological interpretability to the imaging-derived signatures, thereby offering insights into the underlying tumor biology. It serves to bridge imaging phenotypes with molecular mechanisms. We agree with the reviewer that multi-modality integration holds great potential, and we will consider such approaches in future studies when addressing clinical questions where complementary preoperative data are available and could further improve predictive accuracy. Thank you again for this valuable suggestion.

Comment #6:

Fusion mechanism clarity: In Figure 1, the fusion mechanism of the two-branch network is not clearly illustrated. Please revise the figure to explicitly describe how tumor and liver features are concatenated or integrated at the decision level.

Response:

We thank the reviewer for this constructive comment. In response, we have revised **Figure 1** and further added more details in the Method parts of the revised manuscript to improve the clarity of the fusion mechanism in our dual-branch network. The updated figure extends the model schematic to more clearly illustrate the separate feature extraction pathways for tumor and liver regions, as well as their integration via cross-attention and concatenation at the decision level. We illustrate the fusion mechanism below.

After the encoder stage, the three-phase features of each branch are concatenated to generate a pancreas tumor feature and a liver feature. Subsequently, each feature is refined through an intra-organ self-attention module to filter out noise and redundancy introduced by multiple downsamplings:

$$F' = \mathrm{SelfAttn}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where F' denotes the enhanced tumor or liver feature, and Q, K, V represent the query, key, and value matrices generated from the tumor/liver features. The refined features are then processed with residual connections and layer normalization:

$$F'' = \mathrm{LN}(F' + F)$$

where LN denotes the layer normalization function, F″ represents the normalized tumor feature (F_t) or liver feature (F_o), and F denotes the original feature. Next, cross-attention is

applied to enable interaction between the tumor and liver features. The feature of tumor branch generates matrices of Q_t',K_t',V_t', while the liver branch generates matrices of Q_o',K_o',V_o'. The cross-attention mechanism is defined as:

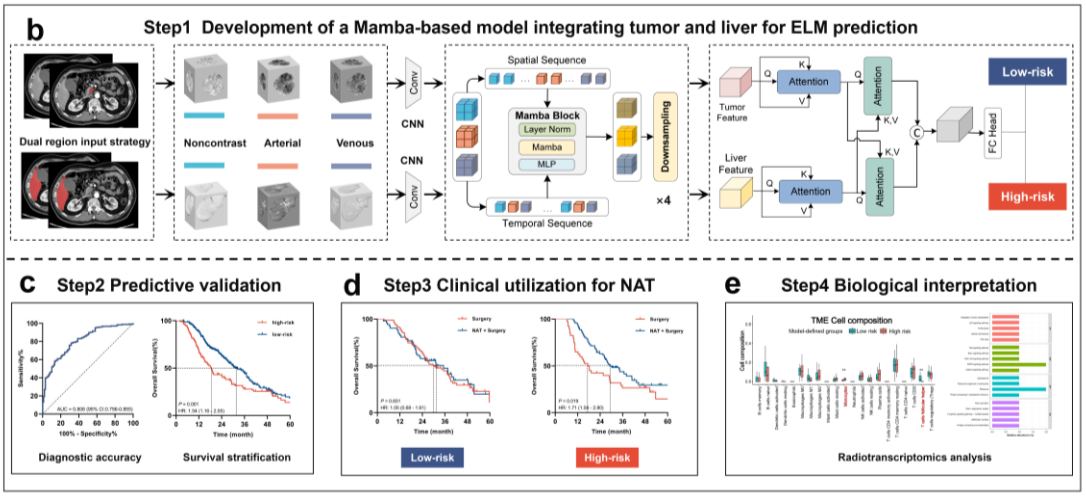
$$F_t^{int} = \mathrm{CrossAttn}(Q'_t \ K'_o \ V'_o),\ F_o^{int} = \mathrm{CrossAttn}(Q'_o \ K'_t \ V'_t),$$

where F_t^{int} and F_o^{int} denote the interaction-enhanced tumor and organ features, respectively. These features are further refined by residual connections and layer normalization:

$$F''_t = \mathrm{LN}(F_t^{int} + F_t),\ F''_o = \mathrm{LN}(F_o^{int} + F_o)$$

Finally, the enhanced tumor and organ features are concatenated along the channel dimension to form the fused representation for final prediction.

This revision, combined with a reorganized layout that aligns with clinical workflow, enhances the overall interpretability of the model architecture. We appreciate the feedback, which has helped us better communicate the design and functionality of our framework.



Comment #7:

Figure 4 labeling: The Kaplan-Meier survival curves are valuable, but the labeling of axes should be clarified to improve interpretability.

Response:

We thank the reviewer for this constructive comment. We agree that the labeling in Figure 4 was insufficiently clear, particularly regarding the subpanel designations (A-D), which were inadvertently omitted in the previous version and may have caused confusion. In response, we have revised Figure 4 with improved axis labels, including explicit definitions of the survival endpoints and time units, to enhance clarity and interpretability. Additionally, we have corrected labeling errors in the related panels of all Figures to ensure consistency and accuracy. We appreciate this valuable feedback.

Comment #8:

The Github repository provides detailed information about the proposed model, including Instructions about preprocessing, training, and inference.

Response:

We thank the reviewer for acknowledging the transparency and reproducibility of our work. Indeed, we have fully documented the model development pipeline in the public GitHub repository, including detailed instructions for data preprocessing, model training, and inference. We hope this facilitates both validation and future adaptation of our approach by the research community.

Response to Reviewer #5:

General Comments:

In this manuscript, Zhao et. al. presents a Mamba-based deep learning model that integrates imaging features from both the pancreatic tumor and the liver to predict occult liver micrometastases (OLM) in PDAC. The model demonstrates robust performance (AUC 0.806 –

0.890) in a large, multi-institutional cohort and provides clinically actionable stratification, as only high-risk patients derived significant survival benefit from neoadjuvant therapy. Radiotranscriptomic validation further supports the biological relevance of the risk groups, strengthening the translational impact of the study. Overall, the work represents a clinically significant advance by offering a noninvasive and generalizable tool for guiding NAT decision-making in PDAC. Its originality and strong clinical applicability make it highly valuable for the field, and the paper is suitable for acceptance after several minor amendments.

Overall Response:

Thank you very much for your positive comments and high recommendations on our study. This study evaluated the utility of Mamba-based predictive model integrating imaging features from pancreatic tumor and the liver to determine the risk of occult liver metastases in a multi-institutional cohort. The model showed good performance in risk stratification and in predicting patient survival. In a subgroup of patients, the radiotranscriptomic analysis provided plausible biologic explanation of the model.

Comment #1:

Provide additional clarification on how Mamba compares to other deep learning architectures (e.g., CNNs, Transformers) for this application in the Introduction.

Response:

Thank you for your insightful comment regarding the superiority of using Mamba. Mamba was selected for its ability to explicitly model spatial and temporal contrast variations through its selective state-space modeling (SSM) mechanism. Unlike CNNs that are limited by local receptive fields or Transformers that incur quadratic complexity, Mamba employs a linear recurrent

formulation to efficiently capture long-range voxel-level dependencies while maintaining scalability to high-resolution 3D data.

This design choice aligns closely with the clinical interpretive process, where radiologists evaluate temporal enhancement patterns across the liver and tumor over multiple phases. By recursively updating hidden states, Mamba achieves fine-grained spatiotemporal sensitivity and global contextual awareness that conventional architectures struggle to balance.

Furthermore, Mamba operates with linear computational complexity, a crucial advantage given our task involves handling large-scale data input--three high-resolution 3D CT volumes across a dual-branch architecture. By efficiently managing this substantial computational burden, Mamba proves exceptionally suited to our task, enabling our method to achieve significantly lower training and inference costs compared to Transformer-based models while maintaining modeling capacity.

In response to the reviewer's suggestion, we have further clarified these advantages in the Introduction of the revised manuscript. Specifically, we now state: "*Mamba dynamically modulates information flow based on input content, enabling effective extraction of spatiotemporal features from sequential imaging data across diverse regions* [26]." We have also added relevant references to support the use of Mamba in medical imaging to strengthen the contextual foundation of our methodological choice.

Reference

26. Qiu, B. *et al.* Multitask deep learning based on longitudinal CT images facilitates prediction of lymph node metastasis and survival in chemotherapy-treated gastric cancer. *Cancer Res* **85**, 2527–2536 (2025).

Comment #2:

Include a more detailed discussion of potential implementation challenges in real-world clinical practice.

Response:

Thank you for this valuable suggestion. We agree that a discussion of real-world implementation challenges is essential for translating predictive models into clinical practice. To better articulate both the clinical utility and translational limitations of our model, we have expanded the Discussion and Limitations sections accordingly.

With regard to clinical value, our findings highlight the potential of the imaging biomarker to support personalized management in pancreatic ductal adenocarcinoma (PDAC). For patients undergoing surgical resection: *Given that early recurrence remains a major contributor to poor outcomes in PDAC, our imaging biomarker demonstrates the potential to non-invasively identify high-risk patients who may benefit from intensified postoperative surveillance and tailored adjuvant therapy [2,22]. Therefore, These findings collectively demonstrate that our model has the potential to extract biologically relevant information from standard imaging phenotypes, thereby guiding clinical decision-making in real-world clinical practice [5,30,31].* In the neoadjuvant setting, current decision-making relies predominantly on anatomical criteria, which fail to capture underlying tumor biology: *Our findings suggest imaging-derived biological risk could refine patient selection for NAT, thereby addressing a key gap in current clinical guidelines and potentially informing future trial design based on biological risk rather than anatomical criteria alone [33,34].*

Nevertheless, this is a retrospective study, and findings require prospective validation in multicenter, diverse populations to assess generalizability and robustness: *Several limitations of this study warrant*

consideration. First, despite the use of propensity score matching and multivariable adjustments, the retrospective study designinherently introduces potential selection bias and unmeasured confounding,which may influence the observed associations. Second, the patient population was derived exclusively from East Asian centers, and the relatively high incidence of ELM in our populationmay limit the generalizability of findings to other ethnic or geographic populations with different clinical characteristics. Third, while our model showed promising results in predicting NAT benefit and stratifying prognosis, direct comparisons between patients who received chemotherapy and those who did not should be interpreted cautiously due to potential confounding. Finally, our findings require prospective validation in large and multicenter cohorts.

Comment #3:

Rephase the abstract using present tense and avoid grammar mistake.

Response:

We sincerely thank the reviewer for the careful reading of our manuscript and for the valuable suggestion. In response, we have carefully revised the abstract to consistently use the present tense and have corrected all grammatical errors. The revised version improves the clarity, accuracy, and overall readability of the abstract.

Comment #4:

Rearrange Fig.1 by shifting the clinical information in 1e to 1a. The layout for original Fig.1a and Extended Fig.1 is redundant.

Response:

We appreciate your valuable feedback on our manuscript. To avoid redundancy and improve clarity, we have revised Figure 1 to provide a clearer representation of the dual-branch Mamba

model, now featuring an enhanced schematic that explicitly illustrates the multi-time-point inputs and cross-attention mechanism. The updated layout also eliminates overlap with Extended Data Figure 1, resulting in a more streamlined and cohesive presentation. Additionally, we have repositioned the clinical information from panel 1e to panel 1a, ensuring a more logical flow of information for readers. The revised Figure 1 is presented below. We believe these modifications we have made several refinements to Figure 1 to enhance its clarity and effectiveness in conveying our research findings. Thank you once again for your constructive comments.

Comment #5:

It is suggested to arrange the subtitle in Result to "Overall study design" rather than "clinicopathological characteristics" for a better summarization.

Response:

We thank the reviewer for the valuable suggestion. We agree that "*Overall study design*" is more appropriate, and we have revised the subsection title accordingly. This change improves the clarity and organization of the Results section.

Second-round review comments

Response to Reviewer #3:

General Comments:

This study developed and validated a model to predict early liver metastases in patients who underwent surgical resection of pancreatic cancer. These results are clinically impactful in triaging patients most likely to benefit from surgery. The authors have adequately addressed the reviewers' comments regarding the clinical and machine learning aspects.

Overall Response:

We sincerely thank the reviewer for the positive evaluation and high recommendation of our study.

Comment #1:

In Extended Data Table 1 and 2: "Differential" should be "Differentiation". There is also a typo in "Lymphvascular invasion" should be "Lymphovascular invasion".

Response:

Thank you for your insightful comment. We have corrected the spelling errors in the tables. "Differential" has been changed to "Differentiation", and "Lymphvascular invasion" has been changed to "Lymphovascular invasion" in both Extended Data Table 1 and Extended Data Table 2.

Response to Reviewer #4:

General Comments:

The authors have made substantial and high-quality revisions to the manuscript. The revised version demonstrates clear improvements in methodological rigor, clarity of presentation, and clinical relevance. The responses are thorough, logically articulated, and well supported by additional analyses. I have no further comments.

Overall Response:

Thank you very much for your thoughtful review; it has truly contributed to the quality of our manuscript.

Response to Reviewer #5:

General Comments:

I appreciate the authors' careful and constructive revisions made in response to the reviewers' comments. The updated manuscript has addressed all points raised in the review, including:

1. Additional clarification in the Introduction regarding Mamba architecture and its comparison to other deep learning approaches (CNNs, Transformers).

2. Expanded discussion on the potential challenges and considerations for real-world clinical implementation.

3. Revision of the abstract to use present tense with improved grammar and clarity.

4. Reorganization of Figure 1 by placing the clinical information from Fig. 1e into Fig. 1a, removing redundancy between the original Fig. 1a and Extended Fig. 1.

5. Adjustment of the Results section subtitle to "Overall study design" for better representation of the content.

The manuscript is now clearer, more informative, and easier to follow. The scientific conclusions are well supported by the data, and the study continues to demonstrate strong clinical significance and potential translational impact. All requested revisions have been satisfactorily completed. I recommend the manuscript for publication in its current form.

Overall Response:

Thank you very much for your positive comments and high recommendations on our study.

Response to Reviewer #6:

General Comments:

This manuscript addresses an important clinical challenge in pancreatic ductal adenocarcinoma, namely the high incidence of early liver metastases following apparently

curative surgery. By integrating multi-phase CT features from both the primary tumor and the liver, the authors propose a deep learning framework to predict early liver metastases and explore whether the resulting risk stratification can identify patients who benefit from neoadjuvant therapy (NAT).

The conceptual motivation for incorporating liver imaging features is sound and distinguishes this work from many prior radiomics and deep learning studies that focus exclusively on the primary tumor. The multi-center design, inclusion of an external validation cohort, and the attempt to assess calibration and clinical utility represent clear strengths. Nevertheless, several major issues require clarification or additional analysis before the conclusions can be fully supported.

Overall Response:

We sincerely thank you for taking the time to thoroughly review our manuscript and for providing such constructive and insightful comments. Your detailed feedback has been invaluable in identifying areas for improvement and enhancing the overall quality of our work. We have carefully considered each of your suggestions and concerns. In response, we have made significant revisions to the manuscript as outlined in our point-by-point response to the reviewers' comments. We believe that these changes have substantially improved the clarity, rigor, and impact of our study.

Comment #1:

Interpretation of early liver metastases (ELM)

The study defines ELM as the occurrence of liver metastasis within six months after surgery. While this endpoint is clinically meaningful, it is not equivalent to direct evidence of pre-existing occult liver micrometastases at the time of imaging. As currently presented, it remains unclear

whether the model is identifying true preoperative micrometastatic disease or a broader phenotype of biologically aggressive tumors that recur early after surgery.

This distinction is important for interpreting the clinical implications of the model. Additional sensitivity analyses using alternative time windows (e.g., shorter or longer postoperative intervals) and a more explicit discussion of this limitation would help clarify what biological process the model is most likely capturing.

Response:

We appreciate the reviewer's valuable insight regarding the interpretation of Early Liver Metastases (ELM). We acknowledge that defining ELM as liver metastasis occurring within six months post-surgery, while clinically relevant as an outcome, does not provide direct evidence of pre-existing occult micrometastases at the time of preoperative imaging. It is indeed challenging to definitively distinguish between these micrometastases based solely on preoperative scans. Therefore, as suggested by reviewer #3, we have updated the terminology throughout the manuscript to consistently use "Early Liver Metastases (ELM)" instead of "occult liver metastases" to better reflect the nature of our endpoint.

As the reviewer mentioned, our deep learning model integrates features from both the primary pancreatic lesion and the liver parenchyma, potentially capturing the biological aggressiveness of the tumor effectively, thereby enabling accurate prediction of post-operative liver metastasis risk. Following the reviewer's suggestion, we conducted additional sensitivity analyses. We evaluated the model's performance in predicting liver metastases occurrence within shorter ($\leqq$ 3 months) and longer ($\leq$ 12 months) post-operative timeframes. These analyses showed consistent trends, with AUC values ranging from 0.735 to 0.856, supporting the model's ability to

stratify risk across different early recurrence windows. And we have added the result in the last of the part of Predictive value of the imaging biomarker for early liver metastases:“ *Beyond the primary endpoint, we also explored the model's predictive capability for liver metastasis occurrence across different time windows. Specifically, the model showed consistent predictive accuracy for both shorter-term (≤ 3 months post-surgery) and longer-term (≤ 12 months post-surgery) liver metastases development (Extended Data Fig. 6).* ”

Comment #2:

Robustness of the neoadjuvant therapy (NAT) benefit analysis

The observation that NAT appears to confer overall survival benefit only in the model-defined high-risk group is intriguing but represents the most methodologically vulnerable aspect of the study. Although propensity score matching is performed, the matching variables appear limited and may not fully capture key clinical factors influencing NAT selection and treatment completion.

Moreover, the retrospective design raises concerns about potential selection bias and immortal time bias, particularly if patients who initiated NAT but did not proceed to surgery were excluded from the analysis. Clarification of patient inclusion criteria and handling of such cases is necessary. To strengthen this claim, the authors should consider expanding the set of covariates used for matching, reporting balance diagnostics, and formally testing treatment-risk interaction effects in a regression framework.

Reesponse：

Thank you for your insightful comments and interest in our exploratory finding. We appreciate your highly professional suggestions regarding the robustness of our analysis. As you

correctly pointed out, this retrospective exploratory analysis faces inherent limitations, including potentially incomplete clinical information and unavoidable selection bias and immortal time bias due to its retrospective nature. As you known we enrolled patients who received NAT and surgery in our study, its hard for us to call back these patients who received NAT treatment but without surgery. We acknowledge these limitations and clearly state them in the Discussion section of the revised manuscript:“ *Third, while our model showed promising results in predicting NAT benefit and stratifying prognosis, we did not include patients who received NAT but did not proceed to surgical intervention. Direct comparisons between patients who received chemotherapy and those who did not should be interpreted cautiously due to potential confounding.*”

Also, in response to the reviewer’s suggestion to expand the set of covariates for matching, we incorporated all available clinical parameters into PSM analysis. Following PSM with these variables, a balanced cohort of 262 patients (131 in each group) was successfully established. We have provided detailed reporting of the covariate balance diagnostics, including both graphical visualizations and table to demonstrate the effectiveness of the matching procedure. Furthermore, the results section of the main text has been updated to reflect these methodological improvements and the resulting balanced cohort characteristics:“ *Propensity score matching (PSM) was performed with potential clinical covariates. After matching, 131 patients were retained in each group with balanced baseline characteristics (Extended Data Table 4 and Extended Data Fig. 7).*”

Based on the latest, most comprehensive matching of pre-operative factors, we have re-conducted the survival analysis:“ *Survival analyses stratified by the model-defined risk groups revealed that high-risk patients who received NAT followed by surgery experienced significantly longer OS compared to those treated with upfront surgery (median OS: 17.4 vs. 34.1 months; HR:*

1.87, 95% CI: 1.27-2.76, P < 0.001) (Fig. 4c). This survival benefit remained significant after PSM (median OS: 18.2 vs. 34.1 months; HR: 1.91, 95% CI: 1.16-3.14, P = 0.004) (Fig. 4d). In contrast, among low-risk patients, NAT did not confer a significant OS advantage before or after matching (all P > 0.05) (Fig. 4a and 4b). "

While it is difficult to include all patients who received chemotherapy without surgery, we also compared survival outcomes for patients who underwent surgery following NAT (with survival time calculated from the start of surgery) against those who had upfront surgery. The results also demonstrated consistent findings:" *The landmark OS analysis, which calculated survival from the time of resection for the NAT group, also supported these findings, showing a consistent survival advantage for high-risk patients receiving NAT (Extended Data Fig. 8).* "

We once again sincerely thank the reviewer for the valuable suggestions. Indeed, obtaining such data is quite challenging, and we have also incorporated as comprehensive an analysis as possible regarding the interesting findings related to the neoadjuvant therapy subgroup within our model. We hope to have the opportunity to conduct prospective cohort studies in the future to provide higher-level evidence in this area.

Comment #3:

Independent contribution of liver-derived imaging features

A central premise of the manuscript is that liver imaging features provide complementary information beyond the primary tumor. While this is supported conceptually by the proposed architecture, more systematic evidence is needed to demonstrate their independent contribution.

In particular, clearer ablation analyses comparing tumor-only, liver-only, and combined tumor–liver models would be helpful. Given the multi-center nature of the dataset, it is also

important to exclude the possibility that liver features are inadvertently capturing center-specific imaging characteristics rather than biologically meaningful signals.

Reesponse：

Thank you for this insightful comment. We acknowledge the importance of providing more systematic evidence for the independent contributions of liver imaging features. To address this concern, we conducted a comprehensive ablation study comparing three distinct models.

Tumor-only model: Only the three-phase pancreatic tumor images were input into the encoder. The cross-attention fusion module was removed, and classification was performed based solely on tumor-derived features.

Liver-only model: Similarly, only three-phase liver images were used as input, without the cross-attention fusion module.

Combined tumor-liver model (combined model): Both branches were jointly trained with the proposed cross-attention-based feature interaction module.

All models were trained and evaluated under identical data splits, preprocessing, and training settings to ensure fair comparison.

The results of this ablation study clearly demonstrate the complementary value of incorporating both tumor and liver features. As shown in the table below, the combined tumor-liver model consistently achieved superior performance compared to the individual component models across both the internal validation and external validation datasets. This substantial improvement confirms that the information derived from the liver parenchyma provides complementary value beyond the primary tumor features, supporting the core premise of our study. To address this, we have supplemented the manuscript with these findings:“ *Albation*

study also demonstrated that the integrated information from both the primary tumor and the liver parenchyma was crucial for optimal predictive performance (Extended Data Fig. 3)."

Additionally, to provide a more comprehensive comparison and highlight the superior performance of our proposed combined model, we have included detailed predictive performance metrics for all three models (Combined, Tumor-only, Liver-only) on the external validation cohort. Crucially, the Combined model demonstrated a significant improvement in sensitivity.

	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Combined	0.806 (0.756-0.855)	0.754 (0.709-0.800)	68.4% (63.4%-73.3%)	75.4% (73.9%-82.6%)
Tumor-only	0.762 (0.711-0.813)	0.696 (0.649-0.743)	49.0% (38.7%-59.5%)	77.9% (72.1%-82.8%)
Liver-only	0.612 (0.548-0.677)	0.556 (0.503-0.608)	56.1% (46.2%-66.3%)	55.3% (49.0%-61.4%)

Comment #4:

Generalizability of the model

Although an external validation cohort is included, all centers are drawn from a relatively homogeneous geographic and ethnic population. This limitation should be more clearly acknowledged when discussing potential clinical deployment. Additional subgroup analyses stratified by scanner vendor or imaging parameters would further support claims of robustness.

Response:

We thank the reviewer for this valuable comment regarding the generalizability of our model. We acknowledge that the patient population in our study was derived exclusively from East Asian centers. We have addressed this limitation explicitly in the Limitations of the revised manuscript:" *Second, the patient population was derived exclusively from East Asian centers and was relatively homogeneous in terms of geography and ethnicity. The relatively high incidence of ELM in our study may limit the generalizability of our findings to other ethnic or geographic*

populations with different baseline clinical characteristics."

To comprehensively evaluate the model's performance across diverse clinical settings, we expanded our subgroup analyses beyond the initial clinical scenario stratifications. Specifically, we performed additional subgroup analyses stratified by scanner vendor to assess the model's robustness under varying technical conditions:" *To further assess generalizability, we conducted subgroup analyses stratified by tumor location, tumor size, and preoperative CA19-9 levels. The model maintained consistently high discriminatory ability across subgroups, with AUCs ranging from 0.821 to 0.872, suggesting its broad clinical applicability. Furthermore, the model demonstrated stable performance with AUCs remaining consistently high with different scanner vendors (Extended Data Fig. 4), suggesting its robustness across varied technical settings.* "

Comment #5:

More detail on the quality control of automated segmentations would improve transparency, including any assessment of inter-reader variability during manual correction.

Response:

Thank you for this important suggestion. We acknowledge that providing more details on the quality control of automated segmentations enhances the transparency and reproducibility of our methodology. As requested, we have included additional information regarding the assessment of inter-reader variability during the manual refinement process.

Specifically, to evaluate the consistency and accuracy of the manual corrections applied to the automated segmentations, we randomly selected 100 patients. Two radiologists independently reviewed and refined the segmentations for the two key volumes of interest: the pancreatic cancer lesion and the whole liver. The Dice Similarity Coefficient (DSC) was calculated to quantify the

overlap between the segmentations produced by the two radiologists. The resulting average DSCs were 0.94 for the pancreatic cancer lesions and 0.99 for the liver segmentations, indicating excellent inter-reader agreement and high-quality manual refinements guided by the automated model.

We have incorporated this information into the Methods section under the subsection detailing the segmentation procedure:" *These masks were subsequently refined manually by radiologists to ensure high accuracy. The inter-rater reliability of these manual refinements yielding Dice Similarity Coefficients (DSCs) of 0.94 for pancreatic lesions and 0.99 for liver segmentations, indicating excellent agreement between readers and validating the quality of the refined masks*."

Comment #6:

The risk threshold is defined using the Youden index; reporting additional clinically interpretable operating points (e.g., fixed sensitivity or negative predictive value) would enhance translational relevance.

Response:

Thank you for this insightful comment regarding the clinical interpretability of our model's risk threshold. Given the clinical priority often placed on high sensitivity to minimize false negatives, we identified an operating point targeting approximately 90% sensitivity. An operating point closest to 90% sensitivity was identified at a threshold of 0.266, yielding a sensitivity of 89.9% (95% CI: 83.4-94.5%) and a specificity of 72.6% (95% CI: 67.4-77.4%). The corresponding positive likelihood ratio (+LR) was 3.28, and the negative likelihood ratio (-LR) was 0.14.

We primarily reported the threshold derived from the Youden index in the main text, as it

represents the optimal balance between sensitivity and specificity. Notably, the Youden index-derived threshold already achieved a high sensitivity of 87.6% (approaching our target of 90%), while maintaining a higher specificity (78.8%) compared to the 90%-sensitivity threshold (72.6%). Therefore, we reported only the threshold derived from the Youden index.

Comment #7:

Calibration performance is mentioned but could be more fully quantified, particularly in the external validation cohort.

Response:

Thank you for this important observation regarding the calibration performance. To address this concern, we have substantially strengthened the calibration analysis. Firstly, we have updated the calibration curves to include the Brier Score (Extended Data Fig. 5). Furthermore, to provide a more rigorous statistical assessment of calibration across validation cohorts, we have included the following results in the main text: " *Moreover, calibration curves showed excellent agreement between predicted probabilities and actual outcomes across all cohorts, indicating strong model reliability. The Hosmer-Lemeshow goodness-of-fit test yielded non-significant P-values in the validation cohorts ($P \geq 0.116$), confirming adequate calibration.*"

Comment #8:

The radiotranscriptomic analysis is based on a relatively small sample size and should be framed more explicitly as exploratory.

Response:

We thank the reviewer for this important observation. We acknowledge that the radiotranscriptomic analysis presented in our study involved a relatively small sample size. As

suggested by the reviewer, we agree that this component of our work should indeed be considered and explicitly framed as exploratory. We revised our manuscript as follows:“ *To elucidate the underlying biological basis of the Mamba-based model in predicting ELM, we performed an exploratory radiotranscriptomics analysis on a subset of patients from the ××× (Center 5) with matched preoperative CECT and bulk RNA sequencing data. For each patient, a risk score was generated using the trained model, and patients were stratified into high- and low-risk groups accordingly.”*

Comment #9:

Additional details on preprocessing and inference pipelines would further improve reproducibility.

Response:

We thank the reviewer for emphasizing the importance of providing additional details on the preprocessing and inference pipelines to enhance reproducibility. We have now clarified these procedures in the revised manuscript.

Preprocessing step：

Briefly, given a 3D abdominal CT volume, preprocessing was performed following the self-configuring pipeline of nnU-Net (v2). Each volume was first read using NibabelIO and reoriented to a consistent axis convention to ensure spatial alignment across cases. A 3D cascade resampling framework was then adopted to address anisotropic voxel spacing. For the full-resolution stage, volumes were resampled to 2.30 × 0.72 × 0.72 mm^3 for pancreatic tumor analysis and 2.50 × 0.73 × 0.73 mm^3 for liver analysis. In the low-resolution stage, the spacing was set to 2.91 × 1.46 × 1.46 mm^3 for pancreatic tumor and 2.50 × 1.28 × 1.28 mm^3 for liver.

Third-order spline interpolation (order = 3) was applied to image intensities, while linear interpolation was used for segmentation masks, with the z-axis handled separately due to anisotropy. After resampling, each volume was cropped to the bounding box of non-zero voxels to remove irrelevant background regions and reduce memory consumption.

Intensity normalization was subsequently performed based on foreground statistics computed from the training set. Voxel intensities were clipped to the 0.5th and 99.5th percentiles of foreground voxels, corresponding to [1, 135] HU for pancreatic tumor and [13, 117] HU for liver in our dataset. Z-score normalization was then applied using the foreground mean and standard deviation (pancreatic tumor: mean 65.12, std 25.51; liver: mean 66.21, std 17.99), which helps mitigate inter-center intensity variability.

In the second stage, a pretrained nnU-Net segmentation model was used to obtain tumor and liver masks. These masks were subsequently refined manually by radiologists to ensure high accuracy. Corrected masks were then multiplied with the preprocessed CT volumes to isolate organ-specific regions. Center-based cropping was then performed to generate fixed-size inputs for classification, with tumor regions cropped to 128 × 128 × 32 and liver regions to 256 × 256 × 64. The three-phase tumor and liver volumes were saved in .npy format for downstream model input.

Inference pipeline：

During inference, each case undergoes the identical preprocessing pipeline described above to ensure consistency with training. The resulting three-phase tumor and liver .npy files are organized according to the directory structure specified in our GitHub repository and fed into the trained dual-branch network to produce binary classification probabilities. For reproducibility, we

provide representative sample data in the data/ directory, including preprocessed .npy files of the non-contrast, arterial, and venous phases for both tumor and liver regions. The data organization during inference follows the same structure as in the training stage, with the three-phase images placed separately in the noncontrast/, arterial/, and venous/ subdirectories. We also release the inference scripts (inference.py and inference_nolabel.py) in the repository to facilitate direct evaluation and deployment. The trained model weights have been already uploaded in the Google Drive (https://drive.google.com/drive/folders/1J9L-5HTu3gXzFG5o3Bs-VoF66-6rph0h). The download link and inference pipeline is also updated in the README file of the repository to enhance reproducibility.

We have added a description of the preprocessing steps in the main text:“ *Given the superior sensitivity of the arterial phase for detecting PDAC and delineating tumor margins, non-contrast and portal venous phase images were co-registered to the arterial phase to enhance spatial consistency. Subsequently, a standardized preprocessing pipeline adapted from the self-configuring nnU-Net (v2) framework was applied. This included reorientation, resampling, and intensity normalization based on foreground statistics derived from the training set. Following preprocessing, automated segmentation of the primary tumor and liver regions were performed using the nnU-Net framework with default parameters 45. These masks were subsequently refined manually by radiologists to ensure high accuracy. The inter-rater reliability of these manual refinements was assessed on a subset of cases, yielding Dice Similarity Coefficients (DSCs) of 0.94 for pancreatic lesions and 0.99 for liver segmentations, indicating excellent agreement between readers and validating the quality of the refined masks.* ” also revised the inference details:“ *During the preprocessing stage, a 3D cascaded resampling framework was adopted to*

address anisotropic voxel spacing. In the full-resolution stage, volumes were resampled to 2.30 × 0.72 × 0.72 mm³ for pancreatic tumor analysis and 2.50 × 0.73 × 0.73 mm³ for liver analysis. In the low-resolution stage, the spacing was set to 2.91 × 1.46 × 1.46 mm³ for pancreatic tumors and 2.50 × 1.28 × 1.28 mm³ for liver. Voxel intensities were clipped to the 0.5th and 99.5th percentiles of foreground voxels, corresponding to [1, 135] HU for pancreatic tumors and [13, 117] HU for liver in our dataset. Subsequently, Z-score normalization was applied using the foreground mean and standard deviation (pancreatic tumor: mean = 65.12, std = 25.51; liver: mean = 66.21, std = 17.99). After segmentation, center-based cropping was performed to generate fixed-size inputs for the classification network. Tumor regions were cropped to 128 × 128 × 32, while liver regions were cropped to 256 × 256 × 64.

During training, several data augmentation strategies were applied to improve model robustness, including random noise injection, random bias field distortion, random flipping, and random motion. The model was optimized using the Adam optimizer with a binary cross-entropy loss function. Training was conducted for 100 epochs with an initial learning rate of 1e−5, which was gradually decreased to 1e−7 using a cosine annealing schedule. All experiments were performed on a single NVIDIA RTX 4090 GPU." Comprehensive implementation details, including code usage and environment setup, are provided in the accompanying README file.

Comment #10:

The authors provide a publicly accessible GitHub repository with source code and a README describing the overall pipeline. The repository is generally well organized and facilitates understanding of the proposed architecture.

However, full reproducibility is currently limited by the absence of example data, pretrained

weights, and detailed instructions for reproducing the reported experiments across cohorts. Clarifying the preprocessing steps, model configuration files, and inference workflow would further enhance the utility of the code as a community resource.

Response:

Thank you for the positive feedback and recognition regarding our code availability. We appreciate your suggestion to enhance reproducibility.

The detailed preprocessing and inference procedures have been described in our response to Comment #9, and the specific details have been incorporated into the README file accompanying the code repository, which describes the overall pipeline. Furthermore, as suggested, we have provided the model configuration files.

We believe these updates address your concerns and enhance the utility of the code as a community resource. If there are any further questions or areas requiring clarification, please do not hesitate to contact us. We are grateful for your valuable suggestions, which have significantly improved the quality of our manuscript.