A deep learning model for personalized intra-arterial therapy planning in unresectable hepatocellular carcinoma: a multicenter retrospective study

Xiaoqi Lin,[a,f] Ran Wei,[b,f] Ziming Xu,[a] Shuiqing Zhuo,[c] Jiaqi Dou,[a] Haozhong Sun,[a] Rui Li,[a] Runyu Yang,[a] Qian Lu,[d] Chao An,[e,**] and Huijun Chen[a,*]

[a]School of Biomedical Engineering, Center for Biomedical Imaging Research, Tsinghua University, Beijing, 100019, China
[b]Department of Gastrointestinal Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, 510080, China
[c]Department of Radiology, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, 510060, China
[d]Tsinghua Changgung Hospital, School of Clinical Medicine, Institute for Precision Medicine, Tsinghua University, Beijing, 100190, China
[e]Department of Minimal Invasive Intervention, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, 510060, China

eClinicalMedicine 2024;75: 102808

Published Online xxx https://doi.org/10.1016/j.eclinm.2024.102808

Summary

Background Unresectable Hepatocellular Carcinoma (uHCC) poses a substantial global health challenge, demanding innovative prognostic and therapeutic planning tools for improved patient management. The predominant treatment strategies include Transarterial chemoembolization (TACE) and hepatic arterial infusion chemotherapy (HAIC).

Methods Between January 2014 and November 2021, a total of 1725 uHCC patients [mean age, 52.8 ± 11.5 years; 1529 males] received preoperative CECT scans and were eligible for TACE or HAIC. Patients were assigned to one of the four cohorts according to their treatment, four transformer models (SELECTION) were trained and validated on each cohort; AUC was used to determine the prognostic performance of the trained models. Patients were stratified into high and low-risk groups based on the survival scores computed by SELECTION. The proposed AI-based treatment decision model (ATOM) utilizes survival scores to further inform final therapeutic recommendation.

Findings In this study, the training and validation sets included 1448 patients, with an additional 277 patients allocated to the external validation sets. The SELECTION model outperformed both clinical models and the ResNet approach in terms of AUC. Specifically, SELECTION-TACE and SELECTION-HAIC achieved AUCs of 0.761 (95% CI, 0.693–0.820) and 0.805 (95% CI, 0.707–0.881) respectively, in predicting ORR in their external validation cohorts. In predicting OS, SELECTION-TC and SELECTION-HC demonstrated AUCs of 0.736 (95% CI, 0.608–0.841) and 0.748 (95% CI, 0.599–0.865) respectively, in their external validation sets. SELECTION-derived survival scores effectively stratified patients into high and low-risk groups, showing significant differences in survival probabilities ($P < 0.05$ across all four cohorts). Additionally, the concordance between ATOM and clinician recommendations was associated with significantly higher response/survival rates in cases of agreement, particularly within the TACE, HAIC, and TC cohorts in the external validation sets ($P < 0.05$).

Interpretation ATOM was proposed based on SELECTION-derived survival scores, emerges as a promising tool to inform the selection among different intra-arterial interventional therapy techniques.

Funding This study received funding from the Beijing Municipal Natural Science Foundation, China (Z190024); the Key Program of the National Natural Science Foundation of China, China (81930119); The Science and Technology Planning Program of Beijing Municipal Science & Technology Commission and Administrative Commission of Zhongguancun Science Park, China (Z231100004823012); Tsinghua University Initiative Scientific Research Program of Precision Medicine, China (10001020108); and Institute for Intelligent Healthcare, Tsinghua University, China (041531001).

Keywords: Hepatocellular carcinoma; Deep-learning; Decision support; Artificial intelligence

*Corresponding author. Department of Medicine, Center for Biomedical Imaging Research, Tsinghua University, Beijing, 100190, China.
**Corresponding author.
E-mail addresses: chenhj_cbir@tsinghua.edu.cn (H. Chen), anchao-1983@163.com (C. An).
[f]Contributed equally to this work.

Research in context

Evidence before this study

Before undertaking this study, we conducted a comprehensive review of existing literature to understand the current state of knowledge on predicting treatment outcomes for unresectable hepatocellular carcinoma (uHCC) using deep learning models. We searched databases covering publications from January 2017 to December 2023. The search terms included "hepatocellular carcinoma," "deep learning," "intra-arterial therapy," "treatment prediction," and "survival outcomes." We included studies that focused on the use of machine learning or deep learning to predict clinical outcomes in uHCC patients, regardless of the language of publication. Studies were excluded if they were reviews, editorials, or case reports.

Added value of this study

Our study contributes significantly to the existing body of evidence by introducing SELECTION, a novel transformer-based deep learning model, which demonstrates high accuracy in prognosticating four treatment plans for uHCC. This study uniquely integrates clinical data with CECT imaging to provide a holistic approach to treatment decision-making. Additionally, the proposed ATOM framework not only enhances the precision of first-line treatment selection but also tailors subsequent combination systemic therapies based on individual patient characteristics. This dual approach of risk stratification and treatment optimization has not been previously explored to this extent in the context of uHCC.

Implications of all the available evidence

The findings from our study, combined with existing evidence, suggest that incorporating advanced deep learning models like SELECTION and ATOM into clinical practice could significantly improve the accuracy of treatment planning for uHCC patients. This approach has the potential to enhance personalized treatment strategies, leading to better patient outcomes. Future research should focus on further validation of these models in diverse patient populations and settings to ensure their generalizability and robustness. Additionally, exploring the integration of these models into routine clinical workflows and evaluating their impact on clinical decision-making and patient outcomes would be valuable.

Introduction

Transarterial chemoembolization (TACE) and hepatic arterial infusion chemotherapy (HAIC) are the two current standard intra-arterial interventional therapy (IAIT) protocols for patients in Asia with unresectable hepatocellular carcinoma (uHCC), which are guided by international guidelines and several randomized clinical trials.[1–3] Moreover, IAITs were administered in combination with molecular-targeted agents (MTAs) or immune checkpoint inhibitors (ICIs), showing significant improvement in long-term survival benefit for uHCC.[4–7]

In the context of hepatocellular carcinoma (HCC), the transition from intermediate to advanced stages often corresponds with heightened tumor invasiveness.[8,9] The intricate landscape of tumor heterogeneity accentuates the complexity surrounding the choice of intervention strategies for patients with uHCC in clinical settings. The therapeutic decision-making process of uHCC remains a subject of ongoing controversy. Critical considerations include imaging features such as tumor burden, the presence of cirrhosis, nonrim hyperenhancement, and infiltrative appearance, all of which play pivotal roles in prognostic predictions for individuals with HCC.[10–13] While these imaging features have gained substantial attention from experts, the core of the decision-making process for IAITs primarily relies on physicians' assessments regarding patient's physical condition and the blood supply status of the targeted HCCs. It is noteworthy that despite the focus on supervisor experience and individual tumor features, the treatment selection and planning for uHCC remain constrained and subject to ambiguity in clinical practice.

Medical artificial intelligence (AI) is advancing from research to clinical use. Despite this, the practical application of image-driven AI for guiding therapeutic decisions in clinical settings remain limited. Dynamic contrast-enhanced computer tomography (CECT) is a standardized imaging technique providing insights into tumor characteristics. We hypothesize a connection between the choice of IAIT and specific imaging features. Previous research showed radiomics features from CECT predicting treatment responses but are less suitable for intermediate-advanced stage HCC.[14–16]

This study aims to develop a deep learning (DL) based decision support system integrating clinical information with CECT to support clinicians in IAIT decision-making for uHCC. To the best of our knowledge, this is the first study using real-world retrospective clinical and CECT data collected to validate the deep learning-based decision support of IAIT schemes in a heterogeneous uHCC population.

Methods

Ethics

This retrospective, multi-institutional study protocol obtained approval from the Institutional Review Board of all participating hospitals (B2022-694) and was conducted following the principles of the 1975 Helsinki Declaration.

Due to retrospective nature of this study, the requirement for written informed consent was waived.

Patient enrolment

Between January 2014 and November 2021, a total of 4773 consecutive patients with uHCC undergoing initial TACE or HAIC were reviewed in 12 tertiary hospitals. Three of the twelve hospitals with 1448 enrolled patients were allocated to contribute to all the train and internal validation sets, attributing to optimization for the proposed model. 277 patients from the other nine hospitals constitute the external validation sets. The distribution of the data source was shown in Supplementary Table S1.1. All of the patients were diagnosed based on the European Association for the Study of Liver (EASL) and the American Association for the Study of Liver Disease (AASLD) guidelines.[1,2] Multiple intrahepatic metastases, distant metastases, or tumour sizes more than 10 cm were considered indicators of uHCC. Further, some suspected cases were confirmed by imaging-guided needle biopsy. The inclusion criteria were as follows: *(a)* age 18–75 years; *(b)* Eastern Cooperative Oncology Group (ECOG) performance status <2; *(c)* Child-Pugh class A or B liver function; *(d)* the management of TACE or HAIC of FOLFOX regimen (oxaliplatin plus fluorouracil and leucovorin) *(e)* the availability of CECT within 2 weeks before the initial treatment; *(f)* Patients with HCC who had a persistent HBV infection were treated with antiviral treatment, and their viral loads were controlled to a low level before receiving the IAIT. The exclusion criteria were as follows: (a) the patients received any therapeutic measures before TACE or HAIC; (b) HCC combined with other malignancies; (c) simultaneous treatment of TACE combined with HAIC; (d) missing image data during perioperative period; (e) lost to follow-up >6 months; (f) uHCC patients receiving simultaneous treatment of TACE combined with HAIC, who have received both TACE and HAIC in uHCC treatment process. CECT scan protocol and the reasons for conducting TACE or HAIC as the first-line therapy were described in Supplementary Information E1. IAIT procedures and the criteria for treatment discontinuation were shown in Supplementary Information E2.1–2.3. The IAIT combination therapy protocols were described in Supplementary Information E2.4.

Study design

The patient enrolment and cohort definitions process can be found in Fig. 1. A total of 1725 uHCC patients [mean age, 52.8 ± 11.5 years; 1529 males] received preoperative CECT scans and were eligible for IAIT. The methodological framework of this study unfolds as follows: initially, the cohort was split into two categories TACE group [(n = 816) and HAIC group (n = 909)] based on interventional modalities. Within these two groups, secondary division was applied based on the receipt of combination systemic therapy, resulting in a subdivision of TACE and HAIC with systemic therapies [TC group (n = 314) and HC group (n = 443)]. The training and internal validation cohorts were chosen from three hospitals with the most enrolled patients. Within these patients, we randomly split them with a ratio of 70/30 for training and internal validation respectively. Where the patients from the remaining nine medical centers were assigned to external validation cohort. The model parameters, meticulously fine-tuned through intensive training with distinctive cohorts outlined in Fig. 1a, prioritizing optimal performance in each internal validation sets. This precise categorization enhances the model's ability to discern nuanced variations in complex patient heterogeneity, contributing to its accuracy.

Data preprocess and automatic delineation

To uphold the uniformity and dependability of both the segmentation algorithm and the envisioned deep learning approach, several pre-processing steps were implemented on the reconstructed CT images. One-dimensional spline interpolation was carried out to standardize the slice thickness to 1 mm across all reconstructed CT images. Additionally, voxel intensities were constrained within −50 to 350 Hounsfield Units (HU) range to null non-liver and non-tumor tissues, thereby enhancing the sensitivity and accuracy of the segmentation, thus subsequent analysis.

Pre-processed CT images underwent automatic segmentation using nnU-Net.[17] nnU-Net is an advanced deep learning architecture designed for semantic image segmentation tasks. Pre-trained nnU-Net models from previous medical image segmentation challenges were made publicly available, therefore segmentation in this task was done by running inference on nnU-Net model pre-trained using dataset from the Liver and Tumour Segmentation (LITS) Challenge. Each of the arterial and portal phase CECT images from individual patients were processed by nnU-Net to provide liver and tumour delineation. For each patient, one informative slice (slice containing the maximum tumour area) was chosen from each of the arterial and portal phase CT images based on the location containing maximum tumour area. The images were cropped to 224 mm * 224 mm using a bounding box covering the whole liver area from the chosen slice. Open-source code and information on preprocessing and inference using nnU-Net can be found (https://github.com/MIC-DKFZ/nnUNet).

Unresectable hepatocellular carcinoma multimodal transformer (SELECTION)

The graphical representation of SELECTION construction is illustrated in Fig. 2a. In this depiction, a transformer-based deep learning model is employed to predict IAIT outcomes, inspired by a recent state-of-the-art proposal for a multi-modal prognostic transformer.[18] Its capability lies in the progressive acquisition of

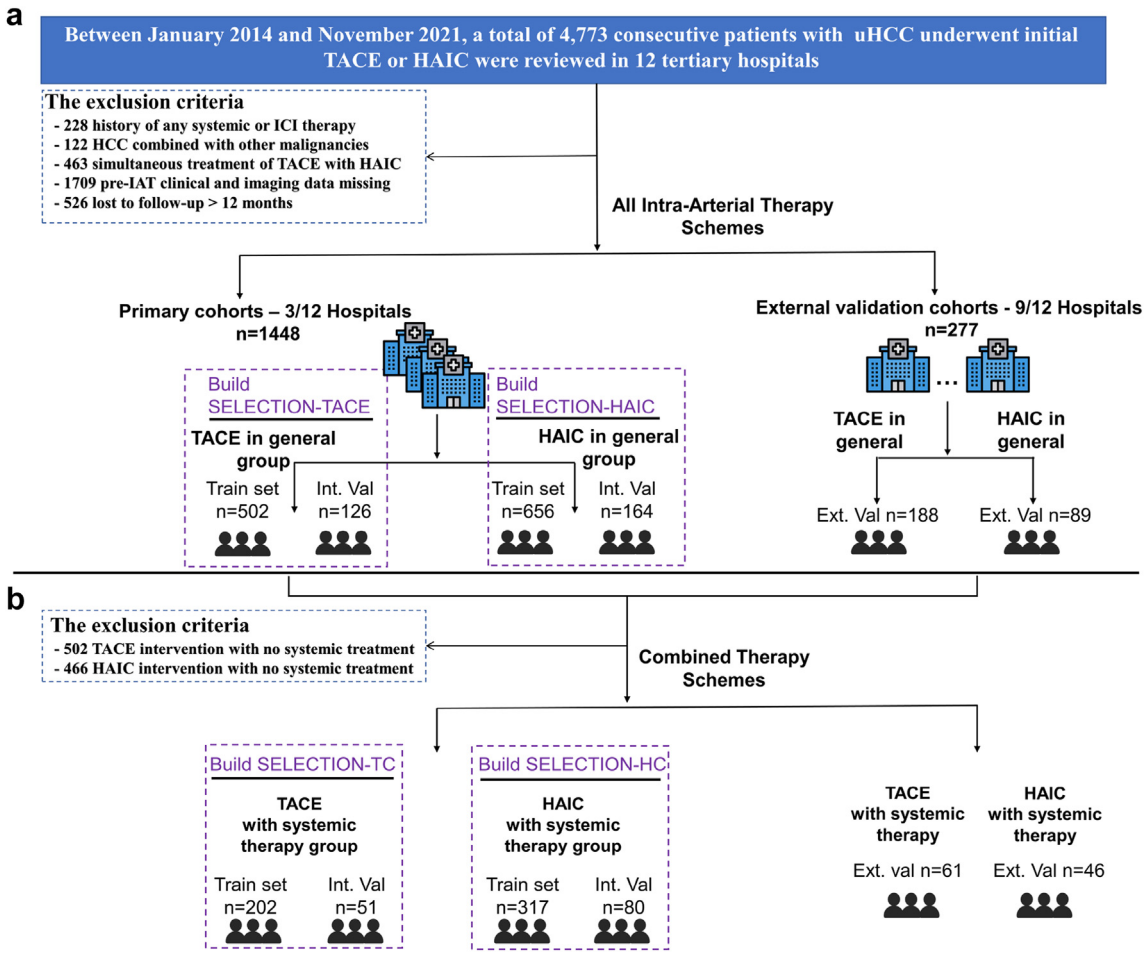

Fig. 1: Patient Enrolment of uHCC patients. (a) Cohort who received TACE or HAIC therapy. (b) Cohort who received TACE or HAIC in combination with systemic therapies, cohort B is considered a subgroup from cohort A. Abbreviations: uHCC, unresectable hepatocellular carcinoma; TACE, Transarterial Chemoembolization; HAIC, Hepatic-arterial infusion chemotherapy.

comprehensive multi-modal representations, treating multimodal input data uniformly as sequences of tokens while integrating the entirety of medical knowledge graph information. SELECTION utilized pre-operative CECT images alongside corresponding demographic baselines and quantitative clinical parameters as its input. Detailed information on the mechanisms of SELECTION, bidirectional multimodal attention and self-attention can be found in Supplementary Material E3.1–E4.2. Performance of SELECTION was compared with a standard Vision Transformer (ViT) utilizing image input only. The detailed information on ViT can be found in Supplementary Materials E3.2. Moreover, we conducted a comparison between SELECTION with a widely employed multi-modal method, a pretrained convolutional neural-network (CNN), ResNet-50, extracted features from CECT images. The extracted features along with clinical variables were then reduced through feature selection, classifiers such as logistic regression predicts the outcome by using the reduced features. In contrast to conventional CNN method, the proposed SELECTION provides outcome prediction by unified processing of multimodal inputs which does not require as much fine-tunning at each stage. Moreover, the multimodal attention used by SELECTION provided multimodal inter-connections at a global level. In comparison, the non-unified methods such as the conventional CNN method only consider the connections between the global representations (i.e., reduced features). This study examined the viability of predicting the overall objective response (OOR) and overall survival (OS) of distinctive uHCC therapeutic plans using SELECTION.

Comprehensive AI-based treatment decision model (ATOM)

The four SELECTION models, cultivated from four distinct cohorts, collectively contribute to a sophisticated

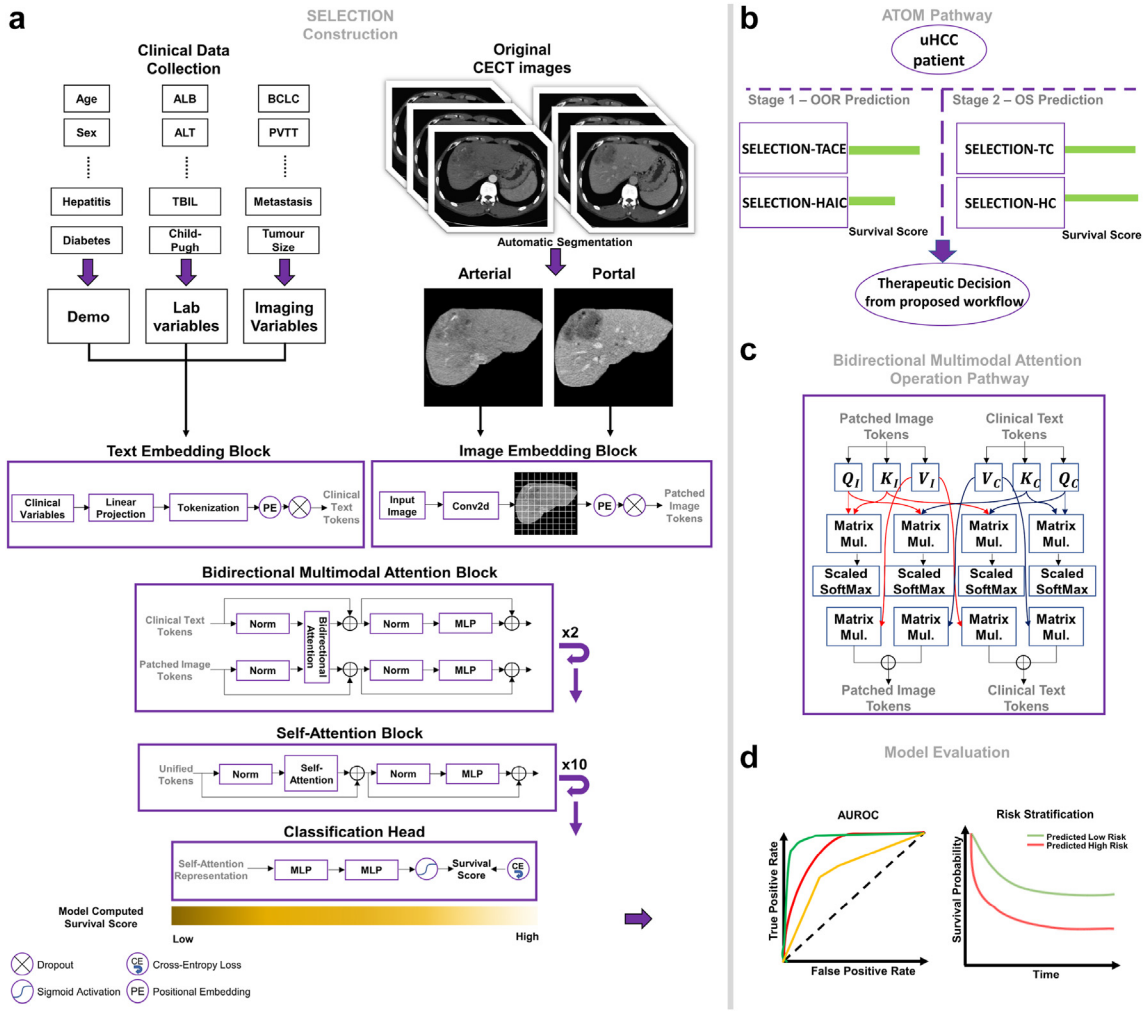

Fig. 2: The workflow of this study, from deep learning model building, decision support pipeline construction to model evaluation. (a) Architecture of SELECTION, Multi-modal input construction was considered from clinical variables and CECT images; (b) Simplified illustration of ATOM in clinical application, where ATOM consist of four SELECTION models trained by four distinctive cohorts, each model will provide a survival score for the patient, through designed decision support pipeline, provide a treatment recommendation according to corresponding SELECTION model. (c) Attention operations in the bidirectional multimodal attention layer. (d) Evaluation methods used to assess the model performance. Abbreviations: uHCC, unresectable hepatocellular carcinoma; MLP, Multilayer Perceptron; Mul., Multiplication; SELECTION, unreSEctabLe hEpatocellular Carcinoma mulTImOdal transformer; AUROC, area under the receiver operating characteristic curve.

decision-making paradigm, enhancing the approach to uHCC management. ATOM describes the comprehensive decision-making process of manipulating the survival scores computed from each SELECTION model, the relationship between ATOM and SELECTIONs is demonstrated in Fig. 2b. ATOM expand its functionality by offering guidance on potential combination therapy alongside the chosen IAIT, introducing a comprehensive consideration of therapeutic strategies. Fig. 3a illustrates the proposed clinical utilization of ATOM. Fig. 3b demonstrated the therapeutic decision-making process of ATOM. The decision-making paradigm is based on making comparisons between patients' survival scores with the median survival scores (MSS) found from each training sets only. It classifies patients into high and low risk groups based on whether the survival score reaches the MSS.

Visualization of SELECTION

Standard approach of attention analysis for vision transformers was used and SELECTION was visualized with GradCam.[19] In the context of multi-head self-attention utilized by SELECTION, attention weights across various heads were averaged from each layer, resulting in an attention matrix. To account for residual connections, an identity matrix was introduced to each attention matrix, followed by normalization of the resultant weight matrices. The weight matrices of different layers from the

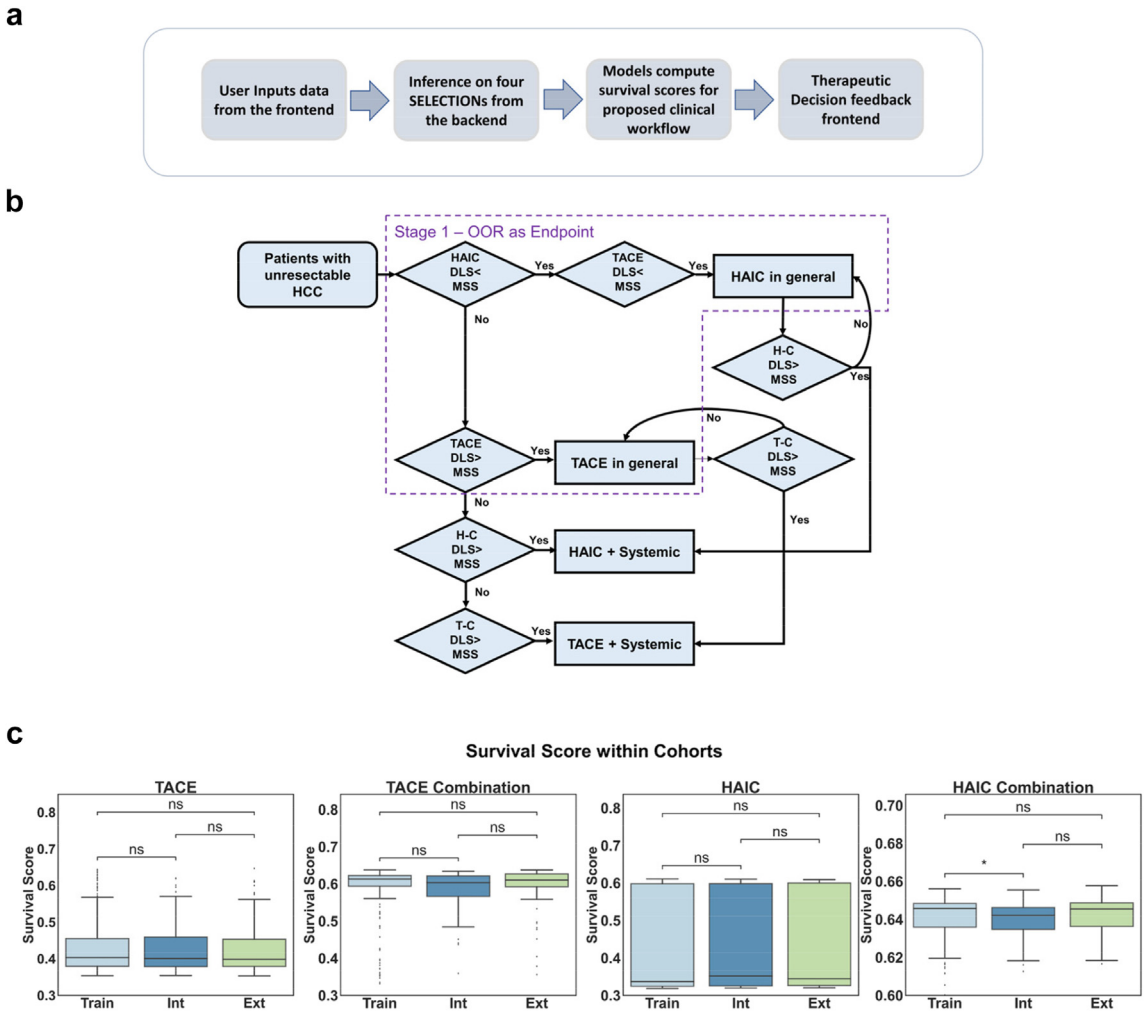

Fig. 3: (a) Overview of ATOM, providing uHCC patient with therapeutic decision planning recommendation. (b) Patient management flowchart demonstrating the risk stratification process, and the ATOM recommendation of therapeutic decision based on survival scores obtained from SELECTION. (c) Survival scores derived by SELECTION found in train, internal validation and external validation in each of the four cohorts in this study. Abbreviation: ATOM, AI-based Treatment Decision Model. P value shows two-sided Mann–Whitney test, Whiskers show minimum and maximum values, boxes represent 25–75% data ranges, ns denotes no significant differences (P > 0.05), * denotes a P value < 0.05. SELECTION: unreSEctabLe hEpatocellular Carcinoma mulTImOdal transformer. DLS, Deep learning score representing the survival score computed by SELECTION in predicting ORR or OS; MSS, median survival score found in the training set; HCC, Hepatocellular Carcinoma; Train, Training datasets; Int, Internal Validation datasets; Ext, External Validation datasets. T-C, TACE with systemic treatment; H-C, HAIC with systemic treatment.

model were multiplied. The culmination of this process yielded an attention map encapsulating the similarity between each input token and the CLS token. Given the role of the CLS token in diagnostic predictions, these similarities served as indicators of the relevance between input tokens and prediction outcomes, facilitating subsequent visualization.

Follow up protocol and endpoints definition

All eligible patients were censored at the last follow-up date [November 30, 2022]. After thorough TACE or HAIC procedure was accomplished, the serum alpha-fetoprotein (AFP) and dynamic contrast enhanced images were examined again at 3–6-month intervals during interventional treatment and at approximately 3-month intervals in the first year and every 6-month intervals follow-up after that. If suspected metastasis was encountered, chest CT, whole-body bone scans, or positron emission tomography-CT were performed selectively.

The primary endpoint was 1-year OS, It was calculated from the date of initial treatment to the following 1-year deadline. The evaluation of the therapy's initial efficacy depends on the ability to differentiate between early responders and non-responders, and 1-year OS is a crucial milestone in this regard. The second endpoint

was overall objective response (OOR) to IAITs based on modified Response Evaluation Criteria in Solid Tumor (mRECIST),[20] including complete response (CR), partial response (PR), stable disease (SD), and progression disease (PD), which was defined as the percentage of patients with CR and PR lasting for over 4 weeks from the first radiological confirmation. 1-year OS serves as a critical milestone for systemic treatments, as it helps to distinguish between early responders and non-responders, which is essential for evaluating the initial effectiveness of the therapy. OOR was chosen to evaluate IAIT outcomes, this ensures effective identification of treatment impacts and to provide metrics consistent with clinical guidelines. We collected a total of 20 clinical variables related to enrolled patient and the variables' definition were described in Supplementary Information E5.

Statistics

Statistical analysis was performed using SPSS version 26.0 (IBM Corp., NY, USA) and RMS package of the R software version 3.5.1 (http://www.r-project.org/). The quantitative variables with mean ± standard deviation or median with interquartile range (IQR) were compared by Mann–Whitney test, Student's t-test, Kruskal–Wallis test and ANOVA. Categorical variables were analyzed using $\chi 2$ tests or Fisher's exact test. The areas under receiver operating characteristic curve (AUC) of different models were compared by Delong test and OS curves of different subgroups were compared using the Kaplan–Meier with log-rank test. All tests of significance were two-sided and a P value < 0.05 was considered statistically significant.

Role of the funding source

This study received funding from the Beijing Municipal Natural Science Foundation (Z190024), the Key Program of the National Natural Science Foundation of China (81930119), The Science and Technology Planning Program of Beijing Municipal Science & Technology Commission and Administrative Commission of Zhongguancun Science Park (Z231100004823012), Tsinghua University Initiative Scientific Research Program of Precision Medicine (10001020108) and Institute for Intelligent Healthcare, Tsinghua University (041531001). The funding was utilized for operation of high-performance computing power, such as graphics cards, patient enrollment, and covering administrative costs. The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Baseline characteristics

Among all the 4773 enrolled patients in this study, 3048 patients were excluded following the exclusion criteria, 816 patients were assigned to TACE group and 909 were assigned to HAIC group. The baseline characteristics stratified by the four therapeutic schemes are shown in Table 1. The distribution of age, ECOG, comorbidity, AFP, tumor size, vascular invasion, metastasis, BCLC stages, albumin, international normalized ratio, platelet, and combination therapeutic schemes including, TKI (Tyrosine Kinase Inhibitor), ICI (Immune Checkpoint Inhibitor) and local treatment, were different significantly between the four groups (all, P < 0.05). Clinical baseline from HCC patients in training set (n = 502), internal validation set (n = 126) and external validation set (n = 188) in TACE group and training set (n = 656), internal validation set (n = 164) and external validation set (n = 89) in HAIC group were outlined in Supplementary Tables S1.2–S1.5. The overall response rate (ORR) was 24.8% (202/816) in TACE group, and 34.2% (311/909) in HAIC group. There was significant statistical difference (P < 0.001) between TACE and HAIC groups. The follow-up period for all uHCC patients receiving IAIT was 13.1 [Interquartile Range: 7.3–24.8] months, and 10.6 [Interquartile Range: 6.8–21.2] months for TACE in general group, Median: 11.9 [Interquartile Range: 6.7–21.2] months for HAIC in general group, Median: 20.9 [Interquartile Range: 9.9–38.1] months for TACE + systemic therapy group and Median: 16.1 [Interquartile Range: 9.7–29.6] months for HAIC + systemic therapy group. The cumulative 1-year, 3-year and 5-year OS among the 4 groups of patients were 1-year: 52.0%, 3-year: 29.5% and 5-year: 20.7% in TACE in general group, 1-year: 56.9%, 3-year: 27.9% and 5-year: 21.0% in HAIC in general group, 1-year: 80.5%, 3-year: 53.6% and 5-year: 41.1% in TACE + systemic therapy group and 1-year: 77.7%, 3-year: 50.5% and 5-year: 40.2% in HAIC + systemic therapy group. For patient undergoing combination systemic therapies, TKI schemes presented median OS of 15.6 months, ICI schemes had median OS of 14.1 months, showing no significant differences between the two systemic treatment schemes (P = 0.736). The ascites, tumor diameter, tumor number, vascular invasion and AFP were independent risk factors for OS in all cohorts (Supplementary Tables S1.13–S1.19).

SELECTION performance for prognosis of distinctive treatment plans

In this study, automatic segmentation results were manually modified by two senior reviewers with more than ten years of expertise in liver imaging and costed around 8 h in total. In total, 3450 CECT slices from the four cohorts were used for development and validation of four individual SELECTIONs (unreSEctabLe hEpatocellular Carcinoma mulTImOdal transformer). Table 2 compared performance between SELECTION with ViT and the multi-modal CNN approach as well as clinical models during training and validation in all four cohorts. Within the deep learning approaches, the CNN approach showed the least regularization abilities,

Baseline Characteristics	Study cohorts				P-value
	TACE	HAIC	TC	HC	
	N = 816	N = 909	N = 314	N = 443	
Age (years), n (%)					
≤65	650 (79.66)	805 (88.56)	253 (80.57)	393 (88.71)	<0.001
>65	166 (20.34)	104 (11.44)	61 (19.43)	50 (11.29)	
Gender, n (%)					
Female	98 (12.01)	98 (10.78)	37 (11.78)	48 (10.84)	0.845
Male	718 (87.99)	811 (89.22)	277 (88.22)	395 (89.16)	
ECOG, n (%)					
0	772 (94.61)	828 (91.09)	294 (93.63)	395 (89.16)	0.002
1	44 (5.39)	81 (8.91)	20 (6.37)	48 (10.84)	
Comorbidity, n (%)					
Absence	755 (92.52)	781 (85.92)	282 (89.81)	382 (86.23)	<0.001
Presence	61 (7.48)	128 (14.08)	32 (10.19)	61 (13.77)	
HBV, n (%)					
Absence	47 (5.76)	60 (6.60)	21 (6.69)	29 (6.55)	0.882
Presence	769 (94.24)	849 (93.40)	293 (93.31)	414 (93.45)	
Ascites, n (%)					
Absence	704 (86.27)	773 (85.04)	282 (89.81)	393 (88.71)	0.090
Presence	112 (13.73)	136 (14.96)	32 (10.19)	50 (11.29)	
Tumor diameter, n (%)					
<5	31 (3.80)	4 (0.44)	18 (5.73)	3 (0.68)	<0.001
[5,10]	410 (50.25)	371 (40.81)	172 (54.78)	188 (42.44)	
>10	375 (45.96)	534 (58.75)	124 (39.49)	252 (56.88)	
Tumor number, n (%)					
≤3	354 (43.38)	332 (36.52)	138 (43.95)	173 (39.05)	0.059
>3	462 (56.62)	577 (63.48)	176 (56.05)	270 (60.95)	
Vascular invasion, n (%)					
Absence	457 (56.00)	268 (29.48)	202 (64.33)	142 (32.05)	<0.001
Presence	359 (44.00)	641 (70.52)	112 (35.67)	301 (67.95)	
Metastasis, n (%)					
Absence	608 (74.51)	542 (59.63)	243 (77.39)	279 (62.98)	<0.001
Presence	208 (25.49)	367 (40.37)	71 (22.61)	164 (37.02)	
BCLC, n (%)					
A	168 (20.59)	91 (10.01)	79 (25.16)	57 (12.87)	<0.001
B	209 (25.61)	143 (15.73)	100 (31.85)	76 (17.16)	
C	439 (53.80)	675 (74.26)	135 (42.99)	310 (69.98)	
ALB (g/L), (mean (SD))	38.73 (5.31)	39.72 (4.62)	39.64 (5.22)	40.23 (4.71)	<0.001
ALT (g/L), (mean (SD))	65.04 (76.51)	62.86 (57.95)	64.30 (78.70)	63.78 (65.96)	0.931
AST (g/L), (mean (SD))	91.11 (126.85)	94.65 (84.66)	85.48 (163.95)	91.42 (86.30)	0.003
TBIL (ng/ml), (mean (SD))	20.40 (27.46)	18.15 (15.20)	19.61 (25.75)	17.77 (16.26)	0.088
PT (ng/ml), (mean (SD))	12.55 (1.74)	12.50 (1.27)	12.45 (1.88)	12.48 (1.27)	0.768
INR (ng/ml), (mean (SD))	1.07 (0.13)	1.09 (0.11)	1.07 (0.16)	1.08 (0.11)	<0.001
PLT (ng/ml), (mean (SD))	215.83 (102.02)	251.28 (121.58)	214.95 (104.38)	253.44 (116.85)	<0.001
AFP (ng/ml)					
≤400	375 (45.96)	353 (38.83)	148 (47.13)	165 (37.25)	0.001
>400	441 (54.04)	556 (61.17)	166 (52.87)	278 (62.75)	
Sequential local treatment, n (%)					
Absence	694 (85.05)	694 (76.35)	193 (61.46)	228 (51.47)	<0.001
Presence	122 (14.95)	215 (23.65)	121 (38.54)	215 (48.53)	
TKI, n (%)					
Absence	605 (74.14)	614 (67.55)	103 (32.80)	148 (33.41)	<0.001
Presence	211 (25.86)	295 (32.45)	211 (67.20)	295 (66.59)	

(Table 1 continues on next page)

Baseline Characteristics	Study cohorts				P-value
	TACE	HAIC	TC	HC	
	N = 816	N = 909	N = 314	N = 443	
(Continued from previous page)					
PD1, n (%)					
Absence	732 (89.71)	723 (79.54)	230 (73.25)	257 (58.01)	<0.001
Presence	84 (10.29)	186 (20.46)	84 (26.75)	186 (41.99)	

Note. -Data are number of patients; data in parentheses are percentage unless otherwise indicated. Mean with standard deviation compared by the ANOVA test. P value < 0.05 suggest statistically significant differences. Abbreviations: HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization; HAIC, hepatic arterial infusion chemotherapy; TC, TAEC used in combination with targeted-immunotheapy; HC, HAIC used in combination with targeted-immunotherapy; ECOG, Eastern Cooperative Oncology Group; HBV, hepatitis type B viral; AFP, α-fetoprotein; ALBI, albumin-bilirubin; ALB, albumin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; PT, prothrombin time; INR, international normalized ratio; TBIL, total bilirubin; PLT, platelet; BCLC, Barcelona Clinic Liver Cancer.

Table 1: **Demographic and clinical characteristics of HCC patients underwent TACE or HAIC.**

having the lowest external validation AUC across TC, HAIC and HC cohorts (TACE: 0.695 [95% CI, 0.621, 0.762], P = 0.44, HAIC: 0.718 [95% CI 0.611, 0.809], P = 0.286, TACE Combination: 0.637 [95% CI, 0.498, 0.761], P = 0.218, HAIC Combination: 0.685 [95% CI, 0.528, 0.817], P = 0.534; significance was tested by the DeLong test with SELECION). Amongst all models, SELECTION though not mostly significant compared to other deep learning approaches, demonstrated the highest performances values. SELECTION-TACE achieved an AUC of 0.788 (95% CI, 0.707, 0.856) and 0.761 (95% CI, 0.693, 0.820) at internal validation and external validation respectively in predicting ORR. SELECTION-HAIC achieved an AUC of 0.854 (95% CI, 0.791, 0.904) and 0.805 (95% CI, 0.707, 0.881) at internal validation and external validation respectively in predicting ORR. The AUC of SELECTION-TC were 0.811 (95% CI, 0.677, 0.907) and 0.736 (95% CI, 0.608, 0.841) in internal and external validation sets respectively in predicting OS. The AUC of SELECTION-HC were 0.837 (95% CI, 0.737, 0.910) and 0.748 (95% CI, 0.599, 0.865) in internal and external validation sets respectively in predicting OS. Results of the DeLong test indicated significant superiority of SELECTIONs over the clinical model-1 in all cohorts expect external validation set from HC cohort (P < 0.05). Clinical Model-2 has performed consistently compared to clinical model-1, where clinical model-2 has outperformed clinical model-1 in TC cohort. Details of clinical models construction can be found in Supplementary Materials E4.3. Furthermore, propensity score matching (PSM) was employed in the external validation cohorts to balance the differences between TACE in general and HAIC in general group, TACE + systemic therapy and HAIC + systemic therapy group to enroll patients with more comparable tumor stage and demographic characteristics (Supplementary Tables S1.7–S1.12). SELECTION consistently showed its' performances in all cohorts before and after PSM and in subgroup analysis (Supplementary Table S1.22). Fig. 3c demonstrated the survival scores derived by SELECTIONs, where no significant differences were seen between the survival scores computed in train, internal validation and external validation sets for HAIC, TACE and TC cohorts (P > 0.05). There were significant differences between the train and internal validation sets for the HAIC combination cohorts (P = 0.019), however no significant differences were present between train and external validation sets. The median survival scores (MSS) found in each train datasets were 0.4035, 0.6130, 0.3368 and 0.6458 for TACE, TC, HAIC and HC cohorts respectively.

SELECTION for identifying survival benefits

Successful stratification of patients into low and high-risk subgroups was achieved by comparing the individual patients' survival scores with the MSS derived from each SELECTION model during the training phase only. The Kaplan–Meier overall survival curves in Fig. 4 illustrated significant differences between the predicted high risk and low risk patients in all four cohorts (P < 0.001). Notably, the best model presented a hazard ratio of 3.2 (95% CI: 2.23, 4.23) in the high-risk group in SELECTION-HC. Higher hazard ratio was recorded in combination therapy cohorts (HC and TC groups) compared to IAIT alone cohorts (HAIC and TACE groups). There was no significant difference between high–risk combination cohort with low-risk IAIT alone cohorts (P = 0.59 and P = 0.10 for TACE and HAIC respectively).

Potential value in ATOM for guidance on recommending treatment

In this study, objective response rate (OOR) as an important indicator was used for evaluating treatment effectiveness of TACE or HAIC. Herein, the value of ATOM (AI-based Treatment Decision Modal) in guiding treatment plan was investigated by analyzing the ORR of uHCC patients who received TACE or HAIC. Moreover, the value of ATOM in recommending additional combination therapy alongside IAIT were evaluated by analyzing the 1-year OS in uHCC patients who received combination therapies.

Models	Cohorts	n	AUC	NPV	PPV	SENS	SPEC	P-value
TACE cohorts								
Clinical model-1	Train	502	0.624 (0.593, 0.655)	0.346	0.872	0.589	0.602	<0.001
	IntVal	126	0.589 (0.513, 0.665)	0.332	0.763	0.547	0.617	<0.001
	ExVal	188	0.578 (0.508, 0.649)	0.289	0.619	0.525	0.592	0.007
Clinical model-2	Train	502	0.624 (0.593, 0.655)	0.346	0.872	0.589	0.602	<0.001
	IntVal	126	0.589 (0.513, 0.665)	0.332	0.763	0.547	0.617	<0.001
	ExVal	188	0.578 (0.508, 0.649)	0.289	0.619	0.525	0.592	0.007
Resnet-50 + LR	Train	502	0.882 (0.850, 0.909)	0.843	0.834	0.434	0.972	0.066
	IntVal	126	0.792 (0.708, 0.861)	0.600	0.798	0.300	0.933	0.892
	ExVal	188	0.695 (0.621, 0.762)	0.875	0.827	0.194	0.993	0.440
ViT	Train	502	0.835 (0.800, 0.866)	0.533	0.915	0.800	0.755	0.924
	IntVal	126	0.755 (0.670, 0.827)	0.476	0.857	0.625	0.766	0.522
	ExVal	188	0.663 (0.590, 0.730)	0.356	0.853	0.525	0.743	0.002
SELECTION	Train	502	0.833 (0.797, 0.865)	0.618	0.897	0.723	0.844	–
	IntVal	126	0.788 (0.707, 0.856)	0.724	0.887	0.656	0.915	–
	ExVal	188	0.761 (0.693, 0.820)	0.456	0.893	0.650	0.791	–
HAIC cohorts								
Clinical model-1	Train	656	0.658 (0.633, 0.683)	0.384	0.779	0.623	0.610	<0.001
	IntVal	164	0.618 (0.545, 0.691)	0.307	0.715	0.536	0.582	<0.001
	ExVal	89	0.588 (0.489, 0.687)	0.392	0.624	0.549	0.538	0.003
Clinical model-2	Train	656	0.565 (0.519, 0.612)	0.626	0.542	0.812	0.314	<0.001
	IntVal	164	0.526 (0.430, 0.622)	0.489	0.583	0.577	0.495	<0.001
	ExVal	89	0.566 (0.444, 0.688)	0.525	0.659	0.489	0.691	0.004
Resnet-50 + LR	Train	656	0.888 (0.861, 0.911)	0.768	0.821	0.578	0.917	0.145
	IntVal	164	0.710 (0.634, 0.779)	0.561	0.758	0.442	0.835	0.003
	ExVal	89	0.718 (0.611, 0.809)	0.730	0.640	0.600	0.762	0.286
ViT	Train	656	0.923 (0.900, 0.942)	0.794	0.936	0.873	0.892	0.653
	IntVal	164	0.848 (0.783, 0.899)	0.623	0.895	0.811	0.766	0.852
	ExVal	89	0.782 (0.681, 0.862)	0.900	0.678	0.587	0.930	0.574
SELECTION	Train	656	0.916 (0.892, 0.936)	0.856	0.924	0.840	0.932	–
	IntVal	164	0.854 (0.791, 0.904)	0.750	0.875	0.736	0.883	–
	ExVal	89	0.805 (0.707, 0.881)	0.875	0.684	0.609	0.907	–
TACE + Systemic cohorts								
Clinical model-1	Train	202	0.618 (0.599, 0.637)	0.287	0.794	0.505	0.619	<0.001
	IntVal	51	0.589 (0.466, 0.712)	0.332	0.707	0.524	0.620	<0.001
	ExVal	61	0.554 (0.422, 0.686)	0.320	0.668	0.519	0.576	0.046
Clinical model-2	Train	202	0.671 (0.578, 0.764)	0.656	0.706	0.832	0.481	0.005
	IntVal	51	0.633 (0.473, 0.793)	0.390	0.576	0.586	0.684	0.105
	ExVal	61	0.639 (0.488, 0.790)	0.574	0.717	0.714	0.577	0.296
Resnet-50 + LR	Train	202	0.974 (0.961, 0.998)	0.829	1.000	1.000	0.400	<0.001
	IntVal	51	0.789 (0.648, 0.894)	0.730	0.818	0.931	0.474	0.738
	ExVal	61	0.637 (0.498, 0.761)	0.614	0.750	0.900	0.346	0.218
ViT	Train	202	0.847 (0.799, 0.901)	0.914	0.645	0.853	0.769	0.519
	IntVal	51	0.810 (0.675, 0.906)	0.880	0.654	0.710	0.850	0.974
	ExVal	61	0.746 (0.620, 0.850)	0.815	0.618	0.629	0.808	0.865
SELECTION	Train	202	0.839 (0.781, 0.887)	0.924	0.600	0.813	0.808	–
	IntVal	51	0.811 (0.677, 0.907)	0.788	0.722	0.839	0.650	–
	ExVal	61	0.736 (0.608, 0.841)	0.800	0.645	0.686	0.769	–
HAIC + Systemic cohorts								
Clinical model-1	Train	317	0.647 (0.616, 0.678)	0.389	0.776	0.593	0.616	<0.001
	IntVal	80	0.621 (0.522, 0.720)	0.315	0.626	0.421	0.578	<0.001
	ExVal	46	0.590 (0.445, 0.735)	0.300	0.619	0.425	0.540	0.215

(Table 2 continues on next page)

Models	Cohorts	n	AUC	NPV	PPV	SENS	SPEC	P-value
(Continued from previous page)								
Clinical model-2	Train	317	0.664 (0.606, 0.722)	0.523	0.728	0.594	0.667	<0.001
	IntVal	80	0.611 (0.474, 0.747)	0.513	0.767	0.516	0.765	<0.001
	ExVal	46	0.736 (0.581, 0.890)	0.608	0.744	0.733	0.621	0.905
Resnet-50 + LR	Train	317	0.929 (0.885, 0.948)	0.817	0.913	0.961	0.656	0.864
	IntVal	80	0.760 (0.650, 0.848)	0.879	0.476	0.823	0.588	0.230
	ExVal	46	0.685 (0.528, 0.817)	0.440	0.789	0.733	0.517	0.534
ViT	Train	317	0.884 (0.844, 0.917)	0.879	0.807	0.883	0.800	0.073
	IntVal	80	0.808 (0.705, 0.888)	0.925	0.519	0.790	0.778	0.678
	ExVal	46	0.690 (0.537, 0.818)	0.474	0.778	0.600	0.677	0.532
SELECTION	Train	317	0.925 (0.890, 0.952)	0.873	0.886	0.939	0.775	–
	IntVal	80	0.837 (0.737, 0.910)	0.938	0.469	0.726	0.833	–
	ExVal	46	0.748 (0.599, 0.865)	0.529	0.793	0.600	0.742	–

Note. —Significance of AUC was tested by the DeLong test between the other models and SELECTION. Abbreviation: SELECTION, unresectable hepatocellular carcinoma multimodal transformer; ViT, Vision Transformer; AUC, area under receiver operating characteristic curve; ACC, accuracy; SENS, sensitivity; SPEC, specificity; PPV, positive predictive value; NPV, negative predictive value; LR, Logistic regression classifier; TACE, transarterial chemoembolization; HAIC, hepatic arterial infusion chemotherapy; TC, TAEC used in combination with targeted-immunotheapy; HC, HAIC used in combination with targeted-immunotherapy; IntVal, Internal validation cohort; ExVal, External validation cohorts; PSM, Propensity score matching.

Table 2: **The performance comparison of different models without PSM.**

Table 3 tabulated the ORR and 1-year OS rates from the external validation cohorts. It depicted variations in ORR/OS based on whether the actual treatment aligned with the ATOM's recommendation. For both TACE and HAIC groups, the ORR was significantly higher ($P < 0.001$) when the actual treatment was consistent with ATOM's recommendation. When ATOM recommend against IAIT, significant differences were observed compared to instances where the actual treatment aligns with ATOM recommendations. The TC group exhibited a similar trend, where the OS rate was significantly higher when ATOM recommendation aligned with the actual treatment. However, this pattern was not statistically significant in the HC group, possibly attributable to its smaller sample size.

Interpretation of SELECTION

Four representative patients in the external validation cohorts were selected and the feature maps learned by SELECTIONs were shown in Fig. 5. The intensity of feature maps in the SELECTION-recommended model has shown to be higher than that of other models. Additionally, the attention heatmap from the suggested SELECTION model has shown focus mostly in the tumour area. This is more obvious in heatmap of the portal phase images, potentially due to tumour region having greater contrast during portal phase. Notably, liver parenchyma surrounding the tumour region were often times focused by SELECTION, showing the contribution of liver regions containing tumour-induced systemic environment in the survival prediction process.

Discussion

In recent years, an increasing number of investigations have presented comparisons of efficacy and safety between TACE and HAIC, along with clinical trial findings on IAIT combined with MATs and ICIs. While various models have been developed based on clinical information or radiomics to predict prognostic outcomes after IAIT,[21,22] with applications in HCC including identification of pathological subtypes, and prospective recurrence prediction after various therapeutic schemes using dynamic enhanced imaging, showing promising prognostic performances. However, these models struggle to guide physicians in formulating individualized treatment schemes. This challenge is attributed to small derivation sample sizes, divergent pathologic scoring standards, and limited clinical applicability.

We hypothesized the choice of IAIT decision can be extracted from CECT images and clinical information at the time of uHCC diagnosis with the help of SELECTION and ATOM. SELECTION predicted survival scores supported the clinical outcomes, presenting highest AUC values amongst all the evaluated models (ORR AUC, 0.761 and 0.805 for HAIC and TACE external validation cohorts respectively; OS AUC, 0.736 and 0.748 for TC and HC external validation cohorts respectively). The risk stratification ability of SELECTION was also visible, showing significant differences in survival between stratified high and low risk groups [$P < 0.001$ for all four cohorts].

ATOM, the proposed clinical pipeline ensured objective response to be achieved in order to select the best therapeutic plan for uHCC patients. The ATOM

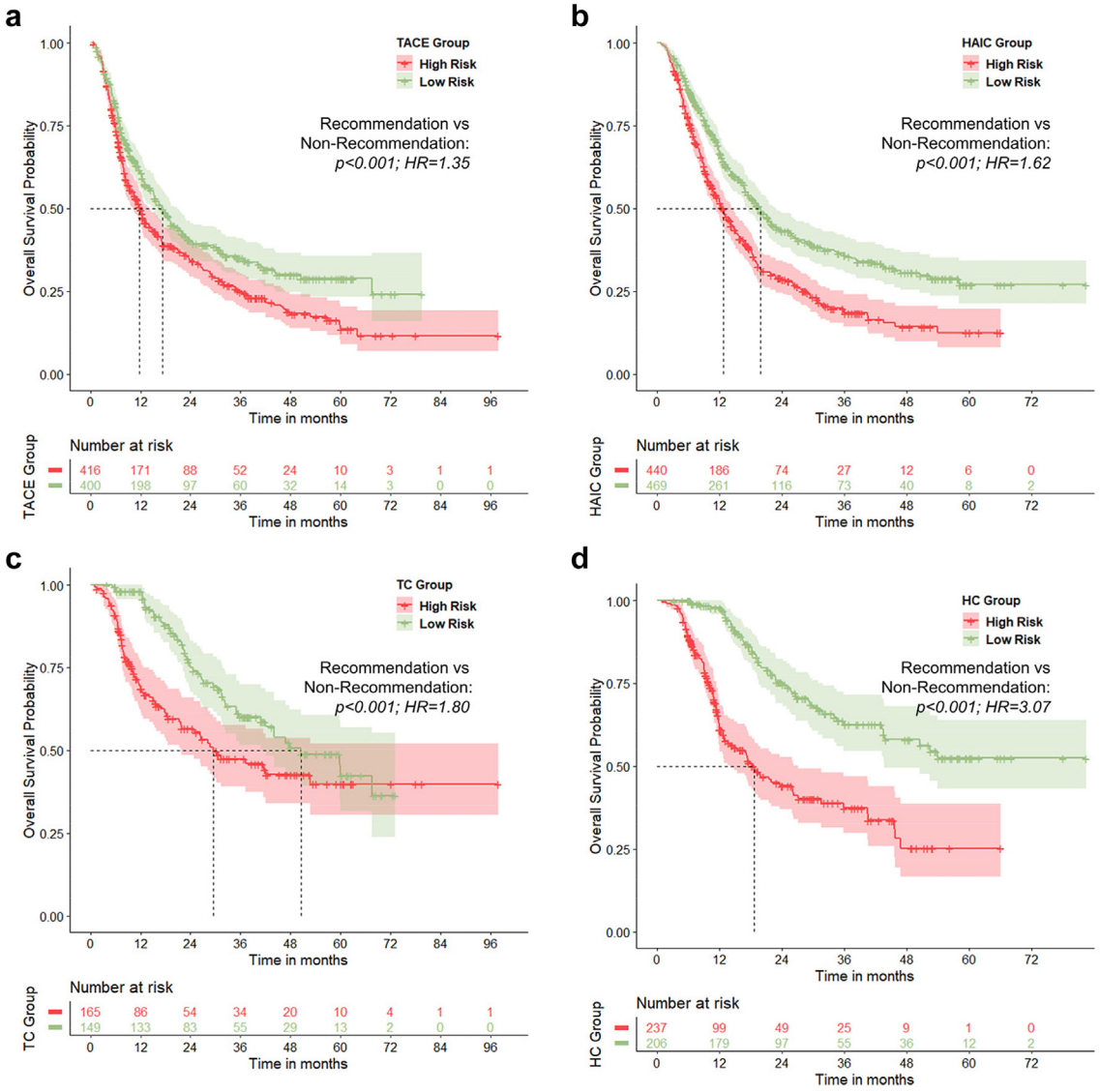

Fig. 4: Kaplan Meier overall survival analysis of high and low-risk groups for (a) TACE group; (b) HAIC group; (c) TC group; (d) HC group. Abbreviation: T-C, TACE with systemic treatment; H-C, HAIC with systemic treatment; HR, hazard ratio.

Stage 1–actual ORR

Cohort	Consistent recommendation	Inconsistent recommendation	P	Recommendation Against IAIT (stage 1)	P	Total number
TACE	30/89 (33.7%)	6/59 (10.2%)	<0.001	4/40 (10.0%)	0.005	40/188 (21.3%)
HAIC	34/49 (69.4%)	3/16 (18.8%)	<0.001	9/24 (37.5%)	0.012	46/89 (51.7%)

Stage 2–Actual OS rate

Cohort	Consistent recommendation	Inconsistent recommendation	P	Recommendation against IAIT (stage 2)	P	Total number
TC group	15/18 (83.3%)	20/38 (52.6%)	0.039	0/5 (0.0%)	0.002	35/61 (57.4%)
HC group	5/11 (45.5%)	9/31 (29.0%)	0.459	1/4 (25.0%)	0.604	15/46 (32.6%)

Note. —Significance was tested by Fisher's exact test. Abbreviation: ATOM, AI-based treatment decision model; ORR, overall objective response; OS, overall survival; Stage 1 refers to making recommendation on TACE or HAIC; Stage 2 refers to making recommendation on combination therapies alongside IAIT.

Table 3: **Actual ORR/OS rate of different cohorts in the external validation set, who has seen consistent recommendation between ATOM and clinician compared with those who had inconsistent recommendations or when no recommendation was suggested by ATOM.**

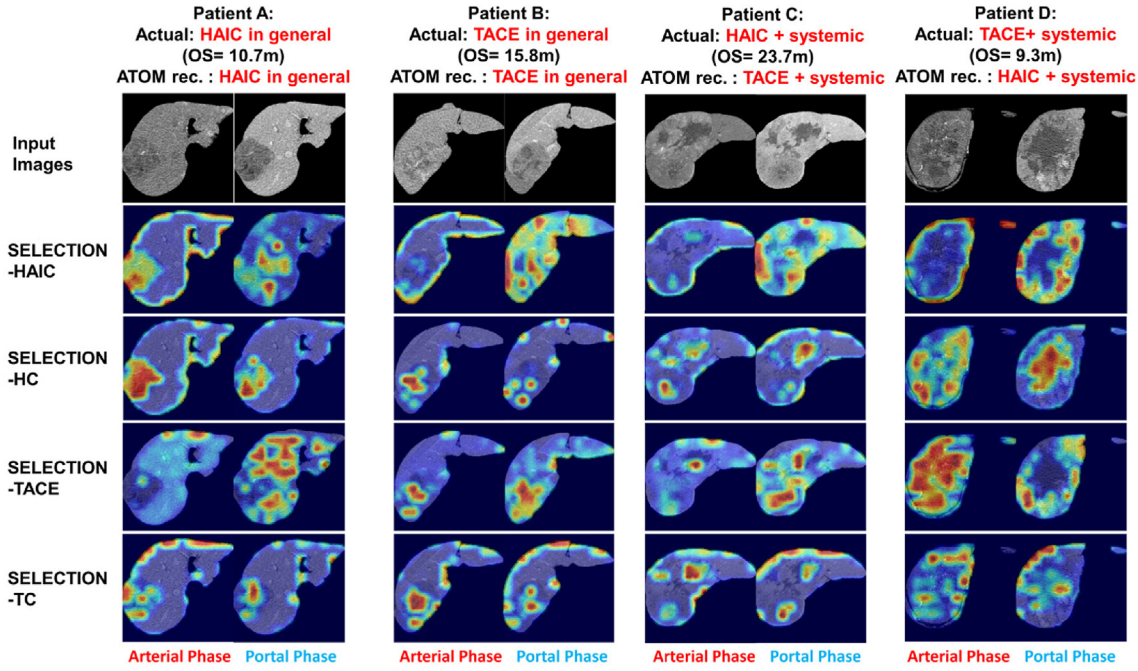

Fig. 5: Feature visualization of SELECTION models on four representative patients using Grad-Cam. Abbreviation: T-C, TACE with systemic treatment; H-C, HAIC with systemic treatment.

framework not only optimizes the selection of first-line treatments but also customizes subsequent combination treatment recommendations based on individual patient characteristics, thereby enhancing the precision and efficacy of therapeutic decisions in uHCC management. It was evidential that when the ATOM recommendation aligns with the actual treatment, there is a significantly higher rate of ORR/OS [$P < 0.05$ for TACE, HAIC and TC cohorts]. Notably, the results revealed extended survival in the IAIT combination groups compared to IAIT alone groups, suggesting that, for high-risk uHCC patients, TACE or HAIC combination therapies should be the recommended first-line therapeutic approach. Our findings align with BCLC guidelines and are consistent with prior research, where after stratification, low-risk group in combination cohorts has significantly higher survival than that of low-risk group in the IAIT alone cohorts. This agrees with previous findings of combination therapy providing longer OS for uHCC patients.[23–25] Finally, for patients predicted to exhibited OOR to initial IAITs only, our findings would be able to caution against anticipating subsequent targeted-immunotherapy or molecular targeted agent to avoid potential economic loss and complications. For SELECTION-high risk IAIT patients and those SELECTION recommending against IAIT altogether patients, proactive treatment scheme adjustments or post-operative prevention and monitoring strategies were to be implemented, emphasizing the need for intensive surveillance and adjuvant systemic therapies due to the elevated risk of death.

Prior investigations predominantly focused on a singular treatment modality recommendation, lacking the capacity to address the fundamental challenges of delivering optimal decision support. Recent studies have been looking into treatment decision between two treatment options for HCC, by forecasting recurrence probabilities of HCC patients and choosing the treatment with higher probabilities.[26,27] Unlike previous proposed models and pipelines, ATOM swiftly and precisely performs risk stratification for uHCC patients undergoing four distinctive IAIT schemes considering multimodal input data. Furthermore, unlike previous studies, ATOM was able to provide meaningful recommendations against IAIT altogether, preventing unnecessary complications.

In the past, studies have often overlooked the intricate interconnection between clinical variables and medical images. Most commonly used multi-modal approach, combining the feature extraction ability of CNN such as ResNet-50 with classification ability of classifiers such as logistic regression, has shown excellent prognostic performances, the generalizability of such method however has yet to be extensively studied.[28,29] Overfitting concern was observed in our application of the CNN multi-modal approach, where comparable performance was noted between the traditional late-fusion CNN approach and

the proposed SELECTION in all internal validation cohorts, but lower performance was evident in external validation cohorts. Besides, the results comparison between ViT and SELECTION demonstrated the importance of clinical features in prognostication performance. Although not statistically significant for most cohorts, higher AUC values were presented by SELECTION in almost all the cohorts when compared to ViT.

The analysis of attention map features learned by SELECTION highlighted the potential roles of tumour micro-environment, in addition the surrounding tumour-induced systemic environment in the liver parenchyma. The intensity displayed in the heatmap reflected heterogeneity, with variations in intensity among different SELECTION models for the same patient indicating distinct focal points elucidated by SELECTION trained on different IAIT schemes. Distinctions in feature maps between arterial and portal phase images were observed, indicating varying concentrations of contrast agents in the liver vessels, potentially influenced by tumor microvessels and the tumor microenvironment. Previous studies exploring mechanisms of conventional chemotherapy resistance had revealed involvement of tumor microenvironment components.[30,31] It seemed to explain the relation between status of tumor micro environment and response to chemotherapy. Based on these studies, we inferred that response to initial IAITs are associated with the tumor microvessels which can be reflected by feature maps. This hypothesis needs further experimental research to explore the mechanism. Visualization of extracted SELECTION features also indicated that response to initial IAITs as well as the following potential combination therapies might be closely related to the area outside the tumor, proving the rationality of the purposed ROI definition. Furthermore, SELECTION exhibits a proclivity for concentrating the liver's edge, a phenomenon accentuated in instances of lower signal-to-noise ratio (SNR) resulting from reduced tube current. We suspect that SELECTION engages in edge tracking of the liver as a response to detecting lower SNR, potentially as a compensatory mechanism for abnormal image parameters. By incorporating survival scores with the MSS calculated from the training dataset, therapeutic decisions were guided by ATOM. While these visualizations provide valuable insights, they should be interpreted with caution, and further studies with larger sample sizes are necessary to validate our findings.

This study is subject to several limitations. Although HAIC is not recommended by most western guideline, HAIC is treated as an effective and safe transcatheter chemotherapy, provides sustained higher concentrations of chemotherapy agents in tumors than intravenous chemotherapy, which has been used for HCC treatment.[6,28] The clinical significance of selection between TACE and HAIC may be more applicable to Asian countries. However, our work remains relevant and significant as it highlights the potential benefits of personalized treatment strategies, which can be adapted to different regional practices and patient populations. The inclusion of patients from 12 hospitals in China introduced variability in treatment practices and procedural techniques across different institutions. Additionally, variations in the selection of TACE drugs could potentially impact the ultimate outcomes. The majority of patients enrolled in our study exhibited large HCC with Hepatitis B Virus infection as the predominant etiology, raising uncertainties about the generalizability of our findings to Western countries where patients often present with a low tumor burden or alcoholic liver cirrhosis as the primary etiology. Another limitation of our study is that while our deep learning model showed higher AUC values compared to other models, these differences were not mostly statistically significant. This suggests potential for improved performance, which could be validated by enrolling patients with more similar baseline characteristics to reduce uncertainty. Moreover, this study did not analyze information on complications during and after IAIT, and administration of TKI, and ICI, necessitating further exploration. Given these limitations, the prospective application of ATOM as a decision support tool in clinical settings requires thorough validation.

In conclusion, SELECTION has successfully predicted uHCC treatment outcomes. ATOM was used to recommend the IAIT scheme in patients with uHCC, it could be easily implemented in clinical practice for physicians and patients to stratify risk and prognosticate quickly and accurately, thereby serving as a more favorable tool to strengthen individualized IAIT schemes.

Contributors

Guarantors of integrity of entire study, X.L, R.W, C.A, H.C.; study concepts/study design, H.C, C.A, X.L, R.W; Data acquisition, data analysis, data interpretation, X.L, R.W, Z.X; manuscript drafting, X.L, C.A; Algorithm optimization, X.L, H.S, R.L, J.D; Verification of underlying data: X.L, R.W,C.A; manuscript revision for important intellectual content, H.C, C.A; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, W.R, X.L, C.A, Z.X, H.C; statistical analysis, W.R, X.L, Z.X,Q.L; and manuscript editing, all authors. All authors read and approved the final version of the manuscript.

Data sharing statement

CT imaging data and clinical information are not publicly available due to patient privacy reasons, they will be made available upon request from the corresponding authors. The code for the model (SELECTION) used in this study can be found on GitHub (https://github.com/banterlin/SELECTION) for test and evaluation purposes.

Declaration of interests

All authors; no relevant relationships. Correspondence and requests for materials should be addressed to X.L, R.W, C.A, H.C.

Acknowledgements
Assistance with the study: none.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102808.

References
1 EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018;69(1):182–236. https://doi.org/10.1016/j.jhep.2018.03.019.
2 Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67(1):358–380. https://doi.org/10.1002/hep.29086.
3 Benson AB, D'Angelica MI, Abrams T, et al. NCCN guidelines® insights: biliary tract cancers, version 2.2023: featured updates to the NCCN guidelines. *J Natl Compr Cancer Netw*. 2023;21(7):694–704. https://doi.org/10.6004/jnccn.2023.0035.
4 Peng Z, Fan W, Zhu B, et al. Lenvatinib combined with transarterial chemoembolization as first-line treatment for advanced hepatocellular carcinoma: a phase III, randomized clinical trial (LAUNCH). *J Clin Oncol*. 2023;41(1):117–127. https://doi.org/10.1200/JCO.22.00392.
5 He M, Li Q, Zou R, et al. Sorafenib plus hepatic arterial infusion of oxaliplatin, fluorouracil, and leucovorin vs sorafenib alone for hepatocellular carcinoma with portal vein invasion: a randomized clinical trial. *JAMA Oncol*. 2019;5(7):953–960. https://doi.org/10.1001/jamaoncol.2019.0250.
6 Ikeda M, Shimizu S, Sato T, et al. Sorafenib plus hepatic arterial infusion chemotherapy with cisplatin versus sorafenib for advanced hepatocellular carcinoma: randomized phase II trial. *Ann Oncol*. 2016;27(11):2090–2096. https://doi.org/10.1093/annonc/mdw323.
7 Zhang T-Q, Geng Z-J, Zuo M-X, et al. Camrelizumab (a PD-1 inhibitor) plus apatinib (an VEGFR-2 inhibitor) and hepatic artery infusion chemotherapy for hepatocellular carcinoma in Barcelona Clinic Liver Cancer stage C (TRIPLET): a phase II study. *Signal Transduct Targeted Ther*. 2023;8(1):413. https://doi.org/10.1038/s41392-023-01663-6.
8 Villanueva A, Hoshida Y, Battiston C, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology*. 2011;140(5):1501–1512.e2. https://doi.org/10.1053/j.gastro.2011.02.006.
9 Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391(10127):1301–1314. https://doi.org/10.1016/s0140-6736(18)30010-2.
10 Wakabayashi T, Ouhmich F, Gonzalez-Cabrera C, et al. Radiomics in hepatocellular carcinoma: a quantitative review. *Hepatol Int*. 2019;13:546–559. https://doi.org/10.1007/s12072-019-09973-0.
11 Qu Q, Liu Z, Lu M, et al. Preoperative Gadoxetic acid-enhanced MRI features for evaluation of vessels encapsulating tumor clusters and microvascular invasion in hepatocellular carcinoma: creating nomograms for risk assessment. *J Magn Reson Imaging*. 2023;60(3):1094-1110. https://doi.org/10.1002/jmri.29187.
12 Martin D, Smet H, Costa ACDS, et al. Tumor burden in patients with early and intermediate-stage hepatocellular carcinoma undergoing liver resection: a retrospective multicenter study on clinical and oncological outcomes. *HPB*. 2023;25(7):836–844. https://doi.org/10.1016/j.hpb.2023.04.001.
13 Jiang H, Yang C, Chen Y, et al. Development of a model including MRI features for predicting advanced-stage recurrence of hepatocellular carcinoma after liver resection. *Radiology*. 2023;309(2):e230527. https://doi.org/10.1148/radiol.230527.
14 Xu X, Zhang H-L, Liu Q-P, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol*. 2019;70(6):1133–1144. https://doi.org/10.1016/j.jhep.2019.02.023.
15 Liu Z, Li Z, Qu J, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. *Clin Cancer Res*. 2019;25(12):3538–3547. https://doi.org/10.1158/1078-0432.CCR-18-3190.
16 Zhang K, Zhang L, Li W-C, et al. Radiomics nomogram for the prediction of microvascular invasion of HCC and patients' benefit from postoperative adjuvant TACE: a multi-center study. *Eur Radiol*. 2023;33(12):8936–8947. https://doi.org/10.1007/s00330-023-09824-5.
17 Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203–211. https://doi.org/10.1038/s41592-020-01008-z.
18 Zhou H-Y, Yu Y, Wang C, et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat Biomed Eng*. 2023;7(6):743–755. https://doi.org/10.1038/s41551-023-01045-x.
19 Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Grad-cam: visual explanations from deep networks via gradient-based localization*. 2017:618–626.
20 Domaratius C, Settmacher U, Malessa C, Teichgräber U. Transarterial chemoembolization with drug-eluting beads in patients with hepatocellular carcinoma: response analysis with mRECIST. *Diagn Interv Radiol*. 2021;27(1):85. https://doi.org/10.5152/dir.2020.19439.
21 Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*. 2020;11(1):1236. https://doi.org/10.1038/s41467-020-15027-z.
22 Wang K, Lu X, Zhou H, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut*. 2019;68(4):729–741. https://doi.org/10.1136/gutjnl-2018-316204.
23 Wang Z, Wang E, Bai W, et al. Exploratory analysis to identify candidates benefitting from combination therapy of transarterial chemoembolization and sorafenib for first-line treatment of unresectable hepatocellular carcinoma: a multicenter retrospective observational study. *Liver Cancer*. 2020;9(3):308–325. https://doi.org/10.1159/000505692.
24 Zhang T, Merle P, Wang H, Zhao H, Kudo M. Combination therapy for advanced hepatocellular carcinoma: do we see the light at the end of the tunnel? *Hepatobiliary Surg Nutr*. 2021;10(2):180. https://doi.org/10.21037/hbsn-2021-7.
25 Voron T, Colussi O, Marcheteau E, et al. VEGF-A modulates expression of inhibitory checkpoints on CD8+ T cells in tumors. *J Exp Med*. 2015;212(2):139–148. https://doi.org/10.1084/jem.20140559.
26 Liu F, Liu D, Wang K, et al. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. *Liver Cancer*. 2020;9(4):397–413. https://doi.org/10.1159/000505694.
27 Ding W, Wang Z, Liu FY, et al. A hybrid machine learning model based on semantic information can optimize treatment decision for naive single 3-5-cm HCC patients. *Liver Cancer*. 2022;11(3):256–267. https://doi.org/10.1159/000522123.
28 Xu Z, An C, Shi F, et al. Automatic prediction of hepatic arterial infusion chemotherapy response in advanced hepatocellular carcinoma with deep learning radiomic nomogram. *Eur Radiol*. 2023;33(12):9038–9051. https://doi.org/10.1007/s00330-023-09953-x.
29 Li Y, Xu Z, Chao AN, Chen H, Li X. Multi-task deep learning approach for simultaneous objective response prediction and tumor segmentation in HCC patients with transarterial chemoembolization. *J Pers Med*. 2022;12(2):248. https://doi.org/10.3390/jpm12020248.
30 Liu L, Zhang R, Deng J, et al. Construction of TME and identification of crosstalk between malignant cells and macrophages by SPP1 in hepatocellular carcinoma. *Cancer Immunol Immunother*. 2022;71(1):121–136. https://doi.org/10.1007/s00262-021-02967-8.
31 Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. 2018;24(5):541–550. https://doi.org/10.1038/s41591-018-0014-x.