npj Digital Medicine

https://doi.org/10.1038/s41746-026-03027-0

Article in Press

Interpretable multimodal deep learning for time-resolved survival prediction after hepatocellular carcinoma resection

Received: 13 February 2026

Accepted: 8 July 2026

Published online: 20 July 2026

Cite this article as: Li, F., Yang, H., Liu, R. *et al.* Interpretable multimodal deep learning for time-resolved survival prediction after hepatocellular carcinoma resection. *npj Digit. Med.* (2026). https://doi.org/10.1038/s41746-026-03027-0

Fan Li, Huanchen Yang, Ruishan Liu, Rundong Wang, Xu Feng, Yangyang Xie, Lian Yang, Wei Zhou, Xia Li, Xiang Wei, Dingbin Yuan, Die Hu, Hao Zhang, Jiahui Zhang, Yuxia Nie, Haibo Qu, Fei Wang, Jing Jia & Gang Ning

Interpretable multimodal deep learning for time-resolved survival prediction after hepatocellular carcinoma resection

Fan Li[1†], Huanchen Yang[4†], Ruishan Liu[1†], Rundong Wang[5†], Xu Feng[8], Yangyang Xie[7], Lian Yang[9], Wei Zhou[10], Xia Li[1], Xiang Wei[1], Dingbin Yuan[1], Die Hu[1], Hao Zhang[1], Jiahui Zhang[1], Yuxia Nie[1], Haibo Qu[3,*], Fei Wang[6,*], Jing Jia[2,*], Gang Ning[3,*]

[1] Department of Radiology, The Third Hospital of Mianyang, Sichuan Mental Health Center, Mianyang, 621000, Sichuan Province, China

[2] Department of Radiology, General Hospital of Ningxia Medical University, Yinchuan 750004, China

[3] Department of Radiology, West China Second Hospital, Sichuan University, Chengdu, 610041, Sichuan Province, China

[4] Department of Radiology, The Third Affiliated Hospital of Shenzhen University (Luohu Hospital Group), Shenzhen 518000, China

[5] Department of General Surgery, The First Affiliated Hospital of USTC, Division of Life Science and Medicine, University of Science and Technology of China, Hefei, Anhui 230001, China

[6] Department of Radiology, Luzhou Hospital of Traditional Chinese Medicine, Luzhou 646000, China

[7] Zhejiang Key Laboratory of Multi-omics Precision Diagnosis and Treatment of Liver Diseases, Department of General Surgery, Sir Run-Run Shaw Hospital, Zhejiang University School of Medicine, 310016, Hangzhou, China

[8] Department of Interventional Medicine Center, The Second People's Hospital of YiBin, Yibin, 644000, Sichuan Province, China

[9] Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

[10] Department of Radiology, Huzhou Central Hospital, Affiliated to Huzhou University, 313000, Huzhou, China

† These authors have contributed equally to this work and share first authorship.

* These authors are co-corresponding authors.

Corresponding author: Gang Ning.

E-mail addresses: lifan1118@126.com (G. Ning)

Address: Department of Radiology, West China Second Hospital, Sichuan University, Chengdu, 610041, Sichuan Province, China.

Abstract

Hepatocellular carcinoma (HCC) exhibits substantial interpatient heterogeneity, leading to markedly variable outcomes and survival even among patients with similar stages and imaging phenotypes. Mainstream staging systems remain suboptimal, whereas pathology-dependent factors and high-cost genomic assays are neither scalable nor timely for clinical decision-making. Existing algorithms provide coarse risk stratification or static binary predictions, failing to capture the time-varying risk of death. We developed and externally validated TEMPO-HCC, a multimodal deep survival model with hierarchical interpretability, to estimate individualized overall survival risk trajectories after curative-intent resection. We curated a six-center cohort of 1,475 patients and integrated multiphasic MRI, postoperative H&E whole-slide images, and perioperative predictors. A discrete-time survival head generated probabilities at 1, 2, 3, and 5 years after surgery. TEMPO-HCC outperformed unimodal models and guideline-based staging systems, achieving C-index of 0.751 and time-dependent AUCs of 0.836, 0.781, 0.812, and 0.680 at 12, 24, 36, and 60 months, respectively, in the external validation cohort. TEMPO-HCC represents a paradigm shift from static binary classification toward clinically actionable temporal risk prediction. Its hierarchical interpretability provides an auditable evidence chain linking macro-scale radiologic phenotypes to micro-scale histopathologic patterns.

Importantly, augmenting guideline staging with TEMPO-HCC improved discrimination, enabling personalized postoperative surveillance and risk-adapted clinical management.

Keywords: Hepatocellular carcinoma; Multimodal deep learning; Multiphasic MRI; Computational pathology; Survival prediction

Introduction

Liver cancer remains a major global health burden, ranking among the leading causes of cancer incidence and mortality[1,2]. Hepatocellular carcinoma (HCC) is the predominant histologic subtype of primary liver cancer, accounting for approximately 75–85% of cases[3,4]. For patients with resectable disease, surgical resection remains a cornerstone curative-intent treatment[5, 6]. However, substantial heterogeneity in tumor biology and the underlying liver microenvironment leads to highly variable postoperative trajectories, and the risks of recurrence and death remain considerable even after curative resection[7, 8]. While genomic testing offers profound insights, its widespread implementation is hindered by excessive costs and logistical delays. In contrast, medical images, including radiologic and histopathologic images, contain abundant high-dimensional visual information that remains underexploited. Many subtle patterns are difficult to discern by visual inspection alone, which may cause radiologists and pathologists to overlook clinically meaningful prognostic or biological cues. An early, accurate, and time-aware prognostic assessment that leverages routinely available perioperative pathology–imaging–clinical information could therefore enable risk-adapted surveillance and recurrence monitoring, and support

rational allocation of adjuvant strategies and clinical trial enrollment—ultimately improving postoperative outcomes and optimizing healthcare resource utilization.

Despite rapid progress in postoperative prognostic modeling for HCC—encompassing conventional staging and clinicopathologic scoring systems (e.g., models integrating tumor burden, liver function, and biomarkers)[9,10], radiomics[11-13], quantitative pathology[14], and deep learning[15, 16]—several critical gaps persist. First, most prior work emphasizes coarse risk stratification (e.g., low/intermediate/high risk) or single time-point binary endpoints (e.g., “recurrence/death within 3 years”), generating relative risks or aggregate scores rather than clinically actionable, time-resolved individualized probabilities.

This limitation restricts precision decision-making in surveillance intensity, adjuvant therapy consideration, and trial eligibility, in which the timing of an event is often as important as its occurrence. Second, many models rely on a single modality (clinical variables, MRI, or pathology), which is insufficient to capture the complex interplay among imaging phenotypes, histologic heterogeneity, and host liver status; even in multimodal settings, fusion is often limited to simple concatenation and underutilizes complementary cross-modal signals. Third, limited interpretability remains a major translational barrier: although deep models may improve performance, they often fail to provide a coherent and traceable evidence chain across pathology regions, imaging substrates, and key clinical factors, undermining clinician trust and adoption. Finally, many studies are

constrained by single-center cohorts or modest sample sizes, leaving generalizability under heterogeneous real-world conditions (MRI protocols, pathology processing, and population characteristics) insufficiently validated. Collectively, these gaps motivate a multi-level explainable deep learning framework that integrates H&E-stained whole-slide images–MRI–clinical information to advance postoperative prognosis from “whether” to “when,” supported by rigorous evaluation in large multicenter cohorts.

Accordingly, as illustrated in **Figure 1**, we propose *TEMPO-HCC* (a ***T***ime-resolved ***E***xplainable ***M***ultimodal ***P***rognostic m***O***del for ***HCC***), a multicenter multimodal deep learning framework that fuses clinical variables, preoperative multiphasic MRI, and postoperative H&E-stained whole-slide images to enable time-resolved survival risk prediction after HCC resection. Moving beyond conventional binary classification and coarse stratification, TEMPO-HCC explicitly targets the transition from “whether” to “when” by producing discrete-time, clinically aligned risk estimates—quantifying the probability of death within multiple postoperative horizons (e.g., 1/2/3/5 years). To enhance clinical utility and credibility, the framework provides multi-level interpretability with a traceable patient-level selection evidence chain spanning (i) contributions of clinical features, (ii) salient MRI regions, and (iii) prognostically informative H&E-stained whole-slide images (WSIs). Importantly, augmenting guideline staging with TEMPO-HCC substantially improved prognostic discrimination. This design aims to deliver a generalizable, transparent, and actionable tool for individualized postoperative management of HCC.

Results

Baseline characteristics

As shown in **Figure 2**, we screened 2,119 consecutive patients with histopathologically confirmed HCC who underwent curative-intent resection and preoperative multiphasic MRI across six institutions (February 2012–January 2024). After excluding patients with prior liver-directed therapy (*n* = 426), poor MRI quality (*n* = 119), or incomplete/unavailable key clinical data (*n* = 99), the final study population comprised 1,475 patients. For model development, patients from Institutions 1–2 (*n* = 702) formed the derivation cohort and were randomly split 8:2 into a training cohort (*n* = 562) and an internal test cohort (*n* = 140). The remaining patients from Institutions 3–6 constituted an **e**xternal validation cohort (*n* = 773). Baseline characteristics were generally comparable across cohorts: most patients were male (training/test/external: 84.9%/89.3%/82.4%) with mean ages of 59.15 ± 11.51, 60.82 ± 10.96, and 58.67 ± 10.66 years in the train, internal test, and external validation cohorts, respectively. Hepatitis B virus (HBV) infection was the predominant etiology (79.5%/76.4%/70.1%) and cirrhosis was frequent (70.8%/73.6%/70.1%). The mean tumor size was approximately 4 cm (training/test/external: 4.31 ± 2.30**,** 4.13 ± 2.38**,** 3.98 ± 1.91 cm). Additional baseline clinicopathologic characteristics are summarized in **Table 1**.

In the derivation cohort, multivariable Cox regression incorporating age, cirrhosis (yes vs no), tumor size, lymph node metastasis (LNMet; yes vs no), AFP category (high

vs low), aspartate aminotransferase (AST), and albumin (ALB) identified four variables as independent predictors of overall survival. Specifically, cirrhosis (HR 1.34, 95% CI 1.07–1.67; $P = 0.011$), larger tumor size (HR 1.17, 95% CI 1.12–1.22; $P < 0.001$), presence of LNMet (HR 1.50, 95% CI 1.07–2.08; $P = 0.017$), and high AFP (HR 1.74, 95% CI 1.41–2.15; $P < 0.001$) were independently associated with increased mortality risk **(Figure 3)**. These independent clinical risk factors were subsequently incorporated into the clinical branch of TEMPO-HCC for model construction.

Performance of predictive models

Accordingly, postoperative H&E-stained whole-slide images, preoperative multiphasic contrast-enhanced MRI, and the independent clinical risk factors identified by multivariable Cox regression were integrated to construct TEMPO-HCC. As shown in **Figure 4** and **Table 2**, TEMPO-HCC demonstrated robust performance for time-resolved OS risk prediction across clinically relevant horizons. In the training cohort, TEMPO-HCC achieved a C-index of 0.845 (95% CI, 0.819–0.873) and an IBS of 0.051 (95% CI, 0.038–0.069), with time-dependent AUCs of 0.983 (95% CI, 0.958–0.996) at 12 months, 0.970 (0.950–0.989) at 24 months, 0.960 (0.938–0.979) at 36 months, and 0.926 (0.885–0.968) at 60 months. In the internal test cohort, the model maintained good discrimination (C-index 0.787, 0.724–0.854) and low prediction error (IBS 0.049, 0.030–0.076), with AUCs of 0.906 (0.819–0.978), 0.845 (0.754–0.910), 0.876 (0.813–0.949), and 0.740 (0.570–0.923) at 12, 24, 36, and 60 months, respectively. Importantly, in the more heterogeneous

external validation cohort, TEMPO-HCC retained consistent generalizability (C-index 0.751, 0.722–0.777; IBS 0.087, 0.076–0.099), yielding time-dependent AUCs of 0.836 (0.728–0.932), 0.781 (0.726–0.827), 0.812 (0.770–0.846), and 0.680 (0.619–0.744) at the corresponding time points. Calibration curves across all cohorts were overall close to the ideal diagonal line, indicating good agreement between predicted and observed OS risks over time. Decision curve analysis (DCA) further demonstrated that TEMPO-HCC consistently achieved higher net benefit than the treat-all and treat-none strategies across clinically relevant threshold probabilities in the training, internal test, and external validation cohorts, with the advantage being most pronounced for 3- and 5-year risk prediction **(Figure S3)**.

Beyond time-specific discrimination and calibration, we next evaluated the clinical utility of TEMPO-HCC for risk stratification. Using the TEMPO-HCC–predicted mortality risk at each postoperative horizon (1, 2, 3, and 5 years), patients were dichotomized into high-risk and low-risk groups, and Kaplan–Meier analyses were performed. As shown in **Figure 5**, TEMPO-HCC consistently yielded pronounced separation of overall survival curves in the training, internal test, and external validation cohorts across all four horizons, with significantly poorer overall survival in the high-risk group compared with the low-risk group and log-rank $P < 0.001$ in each cohort. Notably, survival divergence emerged early after surgery and remained sustained throughout follow-up, indicating that TEMPO-HCC provides stable, time-aware risk stratification that is reproducible under heterogeneous

multicenter conditions.

Comparison and optimization of major guideline-based staging systems

Across both discrimination metrics—C-index and time-dependent AUC—TEMPO-HCC consistently demonstrated superior performance compared to three major guideline-based staging systems (BCLC, CNLC, and AJCC) in the training, internal test, and external validation cohorts. As shown in **Figure 6 and Table S3**, in the external validation cohort, TEMPO-HCC achieved a significantly higher C-index than BCLC, CNLC, and AJCC (0.751 vs. 0.598, 0.633, and 0.561, respectively). TEMPO-HCC also exhibited markedly better time-dependent AUCs at 1-, 2-, 3-, and 5-year post-resection. Notably, integrating TEMPO-HCC as an adjunctive component to each staging system led to significant incremental improvements in both the C-index and time-dependent AUC (all $P < 0.001$), details are shown in **Fig. 6j–l and Table S3**. Taken together, these findings underscore that TEMPO-HCC provides substantial incremental prognostic value beyond established guideline methods, significantly refining OS risk estimation for patients with HCC in routine clinical practice.

Ablation study

To systematically quantify the contribution of each modality and to test the hypothesis that multimodal fusion improves prognostic modeling, we performed a comprehensive ablation study comparing the full tri-modal framework with unimodal and bimodal baselines. As shown in **Figure 4 and Table 2**, a clinical-only model trained on

multivariable Cox–selected risk factors (cirrhosis, tumor size, LNMet, and AFP) achieved a C-index of 0.685 (95% CI: 0.648–0.722), reflecting the intrinsic but limited prognostic value of routine clinical variables. A unimodal deep learning model using H&E-stained whole-slide images yielded a C-index of 0.752 (95% CI: 0.720–0.784), indicating that micro-scale tumor morphology provides prognostic signals beyond standard clinical factors. The radiology-only deep learning model, trained on preoperative multiphasic MRI (pre-T1 and contrast-enhanced arterial, portal venous, and delayed phases), delivered the best unimodal performance with a C-index of 0.792 (95% CI: 0.762–0.822), underscoring that multiphasic MRI captures prognostically relevant whole-tumor phenotypes and spatiotemporal heterogeneity across dynamic enhancement phases. Integrating radiology and WSI further improved discrimination (C-index 0.818), suggesting complementary and non-redundant information across macro- and micro-scale representations. Finally, the fully integrated tri-modal TEMPO-HCC achieved the highest performance (C-index 0.845, 95% CI: 0.819–0.873), significantly outperforming unimodal and bimodal models ($P < 0.05$), with consistent gains across time-dependent AUCs. Collectively, these results support that each modality contributes unique prognostic information and that the proposed deep multimodal framework effectively fuses heterogeneous data streams into a cohesive and robust survival signature.

Subgroup analyses

Subgroup analyses were performed across the multicenter external validation cohort to

evaluate the stability of TEMPO-HCC. The C-indices for centers 3–6 were 0.755 (0.710–0.800), 0.712 (0.635–0.789), 0.745 (0.698–0.792), and 0.765 (0.730–0.800), respectively. Model performance remained largely consistent across centers, with a modest decrease observed in Center 4 (n = 104); time-dependent AUCs showed a similar pattern **(Figure S2a–c and Table S4)**. We hypothesize that the slight performance drops in center 4 may be attributable to its relatively limited sample size. To evaluate whether the predictive performance of TEMPO-HCC was influenced by key perioperative clinical characteristics, we further performed stratified analyses by tumor size, AFP level, and cirrhosis status in the external validation cohort. These analyses demonstrated stable C-indices and time-dependent AUCs, with no statistically significant differences across subgroups **(Figure S2d–l and Table S4)**. Notably, these data were derived from collaborating hospitals with heterogeneous imaging protocols and patient demographics. The consistent performance across diverse external datasets highlights the robustness and generalizability of TEMPO-HCC, suggesting its ability to learn fundamental biological and clinical prognostic patterns that are not specific to the training institution.

Multi-level interpretability with clinical insights

As shown in **Figure 7**, TEMPO-HCC provides multi-level interpretability spanning clinical features, MRI, and histopathology. At the clinical level **(Figure 7a and b)**, global and patient-specific SHAP analyses consistently highlighted AFP, tumor size, cirrhosis, and LNMet as the principal drivers of time-resolved mortality risk, with higher AFP and larger

tumors contributing to higher predicted risk. At the imaging level, Grad-CAM maps on multiphasic MRI predominantly localized model attention to irregular tumor margins and heterogeneously enhancing regions across phases—features that are commonly associated with poor prognosis in patients with hepatocellular carcinoma **(Figure 7c–f)**. At the pathology level, the Grad-CAM heatmaps exhibited strong and consistent spatial concordance between the highlighted “hot spots” and the true location and extent of pathologic lesions, while effectively suppressing attention to surrounding normal tissue **(Figure 7g and h)**. Together, these explanations form a coherent evidence chain linking clinical risk factors to lesion-centric imaging and tissue substrates, supporting the model’s transparency and clinical credibility.

Discussion

In this multicenter study, we developed and externally validated TEMPO-HCC, a time-resolved, interpretable multimodal deep learning framework for modeling postoperative survival trajectories in HCC. TEMPO-HCC was designed to address several persistent gaps in postoperative prognostic assessment. Although recent advances in radiomics, computational pathology, and deep learning have shown promise[17-19], many prior models remain constrained by single-modality inputs or shallow fusion, limiting their ability to capture multiscale features associated with postoperative mortality. Moreover, most existing approaches formulate prognosis as a single time-point binary endpoint (e.g., “alive/dead within 3 years”), implicitly discarding the dynamic nature of risk over time—

information that is directly relevant for tailoring surveillance intervals and therapeutic windows. By integrating macroscopic imaging phenotypes, microscopic histologic heterogeneity, and systemic clinical signals into a discrete-time survival formulation, TEMPO-HCC extends prognostic assessment from binary risk stratification to time-specific risk estimation, providing clinically relevant probabilities across multiple postoperative horizons. Furthermore, in routine clinical practice, incorporating TEMPO-HCC as an add-on to major guideline-based staging systems may further improve OS risk prediction for patients with HCC. This integrated approach may play a pivotal role in standardizing postoperative surveillance intensity and schedules, and optimizing the allocation of clinical resources. In this context, TEMPO-HCC should not be interpreted as a strict head-to-head replacement for previously reported HCC deep learning models, because the underlying tasks and evaluation settings are not fully aligned. Rather, its contribution lies in integrating perioperative clinical data, preoperative multiphasic MRI, and postoperative WSIs into a single framework that provides time-resolved survival probabilities, hierarchical interpretability, and independent multicenter validation. From a translational perspective, TEMPO-HCC may be most realistically deployed first in tertiary centers with digital pathology infrastructure, standardized MRI workflows, and basic AI computing support. In such settings, it may function as a background risk-assessment module integrated into multidisciplinary discussion or postoperative follow-up pathways, complementing rather than replacing clinician judgment.

The clinical branch of TEMPO-HCC incorporates four independent predictors of postoperative mortality—tumor size, cirrhosis, LNMet, and AFP—identified by multivariable Cox analysis and consistently highlighted by SHAP. These factors are biologically plausible: larger tumors generally indicate higher burden and a greater propensity for microvascular invasion and early dissemination[20]; elevated AFP often reflects more aggressive, poorly differentiated disease[21, 22]; LNMet signals extrahepatic spread and adverse systemic behavior[23]; and cirrhosis captures limited hepatic reserve and a pro-tumor inflammatory–fibrotic milieu[24,25]. Together, they represent complementary tumor–host risk axes, and the Positive Linear constraint enforces monotonic associations to reduce clinically counterintuitive predictions under real-world heterogeneity.

TEMPO-HCC achieved strong C-index and time-dependent AUC values across cohorts, while maintaining favorable (lower) IBS—performance that is best understood as the product of complementary information content and purpose-built architectural choices rather than any single component. First, the three modalities contribute nonredundant signals across scales: multiphasic MRI captures vascularity and enhancement dynamics (macroscopic phenotype), H&E-stained whole-slide images provides direct evidence of cellular atypia and microenvironmental architecture (microscopic phenotype), and clinical variables represent systemic status and background liver disease (host context)[26,27]. This multiscale complementarity improves ranking performance and time-specific

discrimination, supporting higher C-index and AUC. Second, the MRI branch implements hierarchical attention at both slice and phase levels, enabling the network to (i) emphasize diagnostically informative slices within each phase[28] and (ii) adaptively weight phases (e.g., arterial and portal venous dynamics) to highlight lesion-specific hemodynamics while suppressing background variability. Crucially, the incorporation of a mask-derived tumor coverage signal injects lesion extent explicitly into attention computation, improving robustness against confounding background features and enhancing cross-center stability[29,30]. Third, the pathology branch builds on MIL by adding coordinate-based positional encodings, preserving spatial context that is often prognostically meaningful (e.g., core–margin differences, focal high-grade regions) and reducing the risk of treating whole-slide image patches as an orderless bag[31,32]. Fourth, the fusion module uses nonlinear transformations with regularization to model cross-modal interactions, enabling recognition of high-risk phenotypes that may not be apparent in any single view (e.g., apparently circumscribed imaging morphology coupled with micro-level invasive histology)[33]. Finally, the discrete-time hazard objective with explicit handling of right censoring aligns training with the generative structure of survival data and facilitates well-calibrated survival probability trajectories, consistent with the observed calibration curves, which directly supports reduced overall prediction error (lower IBS) and improved stability of risk estimates across clinically meaningful horizons[34,35]. From an optimization standpoint, initializing transformer backbones with pretrained weights and freezing them during training focuses learning on attention, projection, fusion, and survival heads—an

effective strategy for mitigating overfitting in heterogeneous multicenter settings[36]. Although the relatively high performance in training cohorts may suggest optimistic estimation, model generalizability was evaluated primarily in the internal test and independent multicenter external validation cohorts rather than inferred from training performance alone. The overall stable discrimination, calibration, and IBS across heterogeneous external institutions support robustness beyond the derivation data. This robustness may be partly attributable to our anti-overfitting strategies, including center-level data splitting, frozen pretrained backbones with only lightweight task-specific training, regularized multimodal fusion, and a survival objective tailored for right-censored data.

Beyond discrimination and overall error, TEMPO-HCC demonstrated clinically relevant risk stratification: Kaplan–Meier curves separated early and remained distinct across long-term follow-up. This stability is consistent with the model's design, which learns interval-specific hazards rather than compressing prognosis into a single scalar score. Such "when"-oriented predictions can support more nuanced management: patients at high near-term risk may warrant intensified surveillance and earlier consideration of adjuvant strategies or trial enrollment, whereas sustained low-risk profiles may justify de-escalated monitoring to reduce burden and optimize resource allocation[37,38].

Interpretability was treated as a core translational requirement rather than an afterthought. By aligning clinical SHAP attributions with Grad-CAM saliency on

multiphasic MRI and Grad-CAM–localized H&E regions, TEMPO-HCC provides a coherent, traceable evidence chain from risk factors to lesion-centric imaging and tissue substrates. This design allows clinicians to audit whether high-risk predictions are driven by plausible tumor- and tissue-related cues rather than artifacts or nonlesion correlations, lowering the trust barrier that often limits clinical adoption of deep learning systems.

Several limitations should be acknowledged. First, the retrospective design and absence of real-world deployment testing remain key limitations. Before clinical implementation, TEMPO-HCC must undergo prospective testing in actual workflows to evaluate turnaround time, user acceptance, and its true impact on multidisciplinary decision-making; until then, it remains a promising prognostic framework rather than a deployed decision-support tool. Second, the endpoint was overall survival; extending the framework to recurrence-related endpoints may further align with postoperative management. Third, tumor masks were derived from expert delineations, which may constrain scalability; automated segmentation or weakly supervised localization could facilitate deployment. Fourth, although H&E adds substantial biological depth, integrating molecular or genomic data may further improve identification of aggressive subtypes. Fifth, SHAP and Grad-CAM provide model-faithful associative explanations rather than causal mechanisms, and should be interpreted alongside domain expertise. Importantly, TEMPO-HCC is designed for postoperative prognostic assessment, not for preoperative decision-making or surgical planning. In this context, future refinement through automated

segmentation, lower-burden localization strategies, and reduced dependence on histopathology may further enhance the feasibility and scalability of TEMPO-HCC in real-world clinical workflows. Sixth, this study was conducted in a predominantly HBV-related HCC population, consistent with both the epidemiologic profile of East Asia and the current development trajectory of most medical AI models, which are typically established and validated within single-country or single-region cohorts before broader international evaluation. Nevertheless, this population composition may constrain generalizability, and further validation in geographically and etiologically distinct populations, particularly Western cohorts enriched for HCV-, alcohol-, and NAFLD/NASH-related HCC, is warranted. Seventh, the real-world implementation of TEMPO-HCC may be limited by the requirement for digital pathology, standardized MRI acquisition, and computational support. Prospective studies are needed to evaluate its feasibility and workflow integration in routine practice. Eighth, the noticeably decreased 5-year AUC in external validation likely reflects the inherent temporal uncertainty of long-term outcomes, which are increasingly shaped by dynamic biological heterogeneity and post-recurrence treatments not captured at baseline. Therefore, 5-year risk estimates should be interpreted cautiously, and future studies with longer follow-up are required to stabilize long-term predictions.

In conclusion, we present TEMPO-HCC, a multicenter-developed and externally validated multi-level interpretable multimodal discrete-time survival model that integrates

preoperative multiphasic MRI, postoperative H&E-stained whole-slide images, and clinical risk information to generate time-resolved postoperative mortality risk estimates in resected HCC. By shifting prognosis from binary “whether” predictions to clinically aligned “when” risk probabilities and providing a traceable explanation chain across modalities, TEMPO-HCC—when used as an add-on to major guideline-based staging systems—may further improve OS risk prediction for patients with HCC. Beyond predictive performance, TEMPO-HCC offers a transparent and actionable framework for individualized postoperative risk assessment. Prospective validation and workflow-integrated deployment studies are warranted to determine whether TEMPO-HCC–guided stratification can improve surveillance efficiency, therapeutic decision-making, and patient outcomes in real-world care.

Methods

Study design and participants

This multicenter retrospective study was conducted across six institutions. This study was approved by the Institutional Review Boards (IRB) of The third hospital of mianyang; Sir Run-Run Shaw Hospital, Zhejiang University School of Medicine; Union Hospital, Tongji Medical College, Huazhong University of Science and Technology; Huzhou Central Hospital, Affiliated to Huzhou University; The Second People’s Hospital of YiBin; and The First Affiliated Hospital of USTC, Division of Life Science and Medicine (approval numbers: 2023-179-20, 2024-0323, 2024-0445, 20201022-01, 2024-226-01, and 2025-RE-111). As

only previously acquired, de-identified imaging, pathology, and clinical data were analyzed, the requirement for written informed consent was waived. Consecutive eligible cases were screened and curated at each center according to a harmonized study workflow; details of participating institutions are provided in **Table S1**.

Inclusion criteria were: (1) histopathologic confirmation of HCC on postoperative specimens; (2) curative-intent surgical resection as the primary treatment; and (3) preoperative MRI including multiphasic dynamic contrast-enhanced T1-weighted imaging, comprising pre-contrast T1-weighted imaging (Pre) and post-contrast arterial phase (AP), portal venous phase (VP), and delayed phase (DP), with image data available for subsequent analysis.

Exclusion criteria were: (1) prior liver-directed therapy before MRI and surgery that could substantially alter tumor imaging phenotypes or histologic characteristics (e.g., transarterial chemoembolization, ablation, or radiotherapy); (2) inadequate MRI quality for analysis (e.g., severe artifacts such as motion, missing key phases/sequences, insufficient anatomic coverage, or other non-diagnostic issues); and (3) incomplete key clinical data that precluded model input construction or outcome ascertainment.

Data collection and preprocessing

MRI acquisition and tumor segmentation: All patients underwent preoperative MRI. The four phases (Pre/AP/VP/DP) were used as imaging inputs. To ensure robust imaging analysis, experienced abdominal radiologists delineated HCC lesions on each phase to

generate tumor masks. Annotations were created exclusively on the four phases of multiphasic dynamic contrast-enhanced T1-weighted imaging DCE-T1WI, including Pre and AP, VP, DP phases. For each case, tumor boundaries were delineated slice-by-slice on each phase to obtain phase-specific three-dimensional tumor volumes, leveraging enhancement characteristics and anatomic landmarks to ensure spatially consistent coverage across phases despite temporal variation in contrast uptake. All segmentations underwent standardized quality control, and ambiguous cases or discrepancies were adjudicated through review and consensus to ensure cross-center consistency. For each subject, multiphasic dynamic contrast-enhanced MRI consisted of four phases—Pre and AP, VP, DP phase—with a phase-specific binary tumor mask available for each phase. All images and paired masks were resampled to a unified voxel spacing of 1.3 × 1.3 × 3.5 mm^3 to harmonize geometry across scanners and sites. MRI intensities were standardized to zero mean and unit variance on a per-volume basis; volumes with near-zero intensity variance were set to all zeros to ensure numerical stability. To obtain a consistent 2D representation while preserving tumor coverage, we selected S = 8 axial slices per phase by uniform sampling along the z-axis, centered on the slice exhibiting the largest tumor cross-sectional area. A configurable in-plane transpose (height‑width swap) and optional flips were applied to enforce consistent orientation across scans. Each selected slice was resized to 224 × 224 pixels (bilinear interpolation for MRI; nearest-neighbor for masks), and grayscale MRI slices were replicated to three channels to match ImageNet-pretrained backbones. The resulting input tensors had shapes MRI (B,

4,S,3,224,224) and masks (B,4,S,1,224,224). MRI acquisition parameters are provided in **Table S2**.

Histopathology acquisition and digital processing: Pathology inputs were derived from routine hematoxylin and eosin (H&E)-stained slides of resection specimens, which were subsequently digitized into WSIs. Slides were required to meet basic quality standards (uniform staining, preserved tissue architecture, absence of severe folding or tissue loss). H&E WSIs were processed at ×40 magnification and tessellated into non-overlapping 512×512-pixel patches, with the grid coordinates of each patch recorded to preserve within-slide spatial context. Each patch image was subsequently resized to 224 ×224 pixels to match the input resolution of standard convolutional backbones and color-intensity normalized using the ImageNet mean and standard deviation. For spatial modeling, the patch location was represented by its grid coordinate (x_i, y_i) and normalized to $[0, 1]$on a per-case basis by dividing by the maximum coordinate values observed within the same WSI, as shown in **Equation (1):**

$$\tilde{x}_i = x_i / \max_j x_j,\ \tilde{y}_i = y_i / \max_j y_j.$$

This preprocessing pipeline yielded a set of appearance features and standardized spatial coordinates for each case, enabling downstream learning to jointly leverage histomorphology and relative tumor microenvironment topology. The resulting patch set and normalized coordinates were used as instance-level inputs for subsequent slide-level modeling. Board-certified pathologists confirmed tumor-related regions to support

downstream modeling and interpretability analyses.

Clinical variables: Perioperative clinical variables (baseline characteristics and laboratory tests) were obtained through the electronic medical record (EMR) system of each institution. Independent clinical prognostic factors were first screened using univariable Cox proportional hazards regression, and covariates achieving prespecified significance were subsequently entered into a multivariable Cox model to identify factors independently associated with the time-to-event outcome. The final multivariable Cox model was then used to derive a clinical risk score for each patient, computed as the linear predictor (the weighted sum of selected covariates using the estimated Cox coefficients), yielding a tensor of shape $(B, 1)$ for a batch of size B. To integrate this scalar risk signal into the downstream deep model, the risk score was projected into a low-dimensional embedding space (32 dimensions) via a non-negative linear layer. Specifically, we parameterized the projection matrix with unconstrained raw weights and enforced non-negativity by applying a softplus transform, the non-negative projection matrix and clinical embedding were computed according to **Equation (2):**

$$W = \mathrm{softplus}(W_{\mathrm{raw}}),\ e = r \cdot W + b,$$

where r denotes the clinical risk score, W_{raw} denotes the unconstrained raw weights, W denotes the non-negative projection matrix, and b is a bias term. This construction ensures that the embedding is a monotone non-decreasing function of the clinical risk score along all learned directions, thereby preserving the ordinal risk information while

enabling flexible feature fusion with imaging and/or other modalities.

Outcome definition and follow-up

The primary endpoint was OS, defined as time from the date of surgical resection to death from any cause. OS was selected as the sole study endpoint after careful consideration of endpoint robustness and cross-center consistency, with the detailed rationale provided in **Supplementary Note 1**.

Patients who were alive at the last follow-up or lost to follow-up were censored at the last confirmed date of survival. Follow-up and outcome information were obtained through routine clinical follow-up systems (outpatient/inpatient records and electronic medical record (EMR) system review), supplemented by telephone contact when needed, and recorded using a predefined, uniform data extraction and verification procedure. The follow-up end date was defined by each center; patients without an event by that date were treated as censored.

At all six participating centers, consecutive eligible patients were screened using a harmonized study workflow with predefined inclusion and exclusion criteria. Multi-modal data were retrospectively curated in a de-identified manner according to a standardized cross-center protocol, and outcome/follow-up information was ascertained through routine follow-up systems and medical record review, supplemented by telephone follow-up when necessary, with prespecified procedures for extraction and verification.

To reduce inter-institutional variability, MRI data were uniformly resampled and normalized, WSI data were processed using a standardized patch-based workflow, and multimodal curation/annotation followed harmonized quality-control procedures across centers. Further discussion of heterogeneity handling is provided in **Supplementary Note 2.**

TEMPO-HCC model development

We developed a multimodal, discrete-time survival deep learning framework to estimate patient-specific OS risk after curative-intent resection of HCC, producing time-resolved risk probabilities at clinically relevant horizons (12/24/36/60 months; i.e., 1/2/3/5 years) **(Figure S1)**. TEMPO-HCC integrates: (1) preoperative multiphasic contrast-enhanced MRI (Pre, AP, VP, DP) with radiologist-derived tumor masks, (2) postoperative H&E-stained whole-slide images as variable-length patch sets, and (3) perioperative clinical variables.

MRI branch: For each MRI phase (Pre/AP/VP/DP), we extracted a fixed-length sequence of axial 2D slices sampled along the z-axis with consistent coverage around the tumor region. Slices were resized to 224×224 and encoded using a Swin Transformer backbone pretrained on natural images. To incorporate lesion extent explicitly, slice-level features were augmented with a mask-derived tumor coverage signal. Features were then aggregated via hierarchical attention: (1) slice-level attention within each phase to obtain phase-specific representations, followed by (2) phase-level attention across phases to

yield a compact MRI embedding capturing complementary enhancement dynamics.

Histopathology branch: For postoperative H&E-stained whole-slide images, tissue regions were tiled into patches under standard quality control (e.g., excluding background, blur, and artifacts). Each patch was encoded using a Swin Transformer backbone. To preserve spatial context, a lightweight coordinate-based positional encoding derived from patch locations was injected into patch features. A multiple-instance learning (MIL) attention pooling module then produced a patient-level pathology embedding, enabling the model to focus on prognostically informative histologic patterns despite variable patch counts across patients.

Clinical branch: Perioperative clinical predictors were selected using univariable and multivariable Cox regression in the training cohort only to avoid information leakage. The final multivariable Cox model was used to compute a clinical risk score (linear predictor) for each patient, which was then projected into a low-dimensional clinical embedding via a non-negative linear layer implemented with a softplus transform on the weights.

Multimodal fusion and discrete-time survival head: MRI, pathology, and clinical embeddings were concatenated and processed by a fusion multilayer perceptron (MLP), followed by a discrete-time survival head that output hazard logits over predefined time bins (0–12, 12–24, 24–36, and 36–60 months) aligned with the target horizons. Predicted hazards were converted to survival probabilities and cumulative risks at 1/2/3/5 years, enabling direct “when”-oriented risk reporting rather than single time-point binary

prediction.

Training objective and optimization: Model training minimized the negative log-likelihood of a discrete-time hazard formulation with right-censoring. Optimization used AdamW with mixed-precision GPU training. To reduce overfitting and improve cross-center robustness, Swin backbones were initialized with pretrained weights and kept frozen during training, while optimizing the attention modules, projection layers, fusion network, and survival head. Additional implementation details (preprocessing, hyperparameters, and time-binning strategy) are provided in.

Model interpretability

To enhance transparency and clinical usability, TEMPO-HCC provides multi-level interpretability at the levels of clinical features, MRI, and H&E-stained whole-slide images. For clinical variables, SHapley Additive exPlanations (SHAP) was used to quantify global and patient-specific feature contributions to the predicted time-resolved mortality risk. For MRI, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to generate saliency maps on multiphasic MRI, highlighting influential regions across phases and slices. For H&E-stained whole-slide images, Grad-CAM was used to localize prognostically informative tissue patterns at the patch level, and saliency was aggregated to produce slide-level visualizations.

Statistical analysis

Univariable and multivariable Cox proportional hazards regression analyses were performed to identify independent clinical prognostic factors. Variables that remained statistically significant in the multivariable Cox model (two-sided $P < 0.05$) were considered independent risk factors and incorporated into the clinical branch of TEMPO-HCC.

The prognostic performance of TEMPO-HCC for postoperative survival risk prediction was evaluated at 12, 24, 36, and 60 months. Discrimination was assessed using time-dependent receiver operating characteristic (ROC) analysis with time-specific area under the curve (AUC), together with Harrell's concordance index (C-index). Overall prediction error over time was quantified using the integrated Brier score (IBS). Calibration was examined using calibration curves comparing predicted and observed survival probabilities at each time point. For performance metrics, 95% confidence intervals were estimated by bootstrap resampling. To explore the potential value of TEMPO-HCC in clinical decision-making, DCA was performed by calculating net benefit across a spectrum of threshold probabilities for 1-, 2-, 3-, and 5-year risk prediction in the training, internal test, and external validation cohorts.

For risk stratification, patients were categorized into high- and low-risk groups according to the TEMPO-HCC–predicted risk score (horizon-specific cutoffs were determined in the training cohort via time-dependent ROC analysis). Kaplan–Meier survival curves were generated for the two groups and compared using the log-rank test.

All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant. All analyses were implemented using custom Python scripts.

Data availability

The imaging data used in this study are protected and not publicly available due to personal information protection, patient privacy regulations, and institutional data-sharing policies. However, the data can be made available to qualified researchers for academic purposes upon reasonable request. Interested parties may contact the corresponding author, who will review each request for scientific merit and compliance with data governance requirements.

Code availability

The source code is available online (https://github.com/Fanli8807/TEMPO-HCC).

Acknowledgements

This study was supported by National Key Research and Development Program of China (2024YFC2419303), Science and Technology Department of Sichuan Province (2025YFHZ0068), and Ningxia Natural Science Foundation (2023AAC03684).

Author contributions

Conceptualization, Methodology, Writing-Original Draft: F.L., H.C.Y., R.S.L. and R.D.W.; Data collection: X.L., X.W., D.B.Y., D.H., H.Z., J.H.Z. and Y.X.N.; Software, Resources: H.C.Y., X.F., Y.Y.X., L.Y. and W.Z.; Analysis and interpretation of data: F.L., H.C.Y., R.S.L.

and R.D.W.; Funding acquisition: G.N., H.B.Q. and J.J., Writing-Review & Editing, Supervision: H.B.Q., F.W., J.J. and G.N.

Competing interests

The authors declare no competing financial or non-financial interests.

References

1. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J, Finn RS. Hepatocellular carcinoma. Nature reviews Disease primers 2021;7(1):6.

2. Singal AG, Lampertico P, Nahon P. Epidemiology and surveillance for hepatocellular carcinoma: New trends. J Hepatol 2020;72(2):250-261. doi: 10.1016/j.jhep.2019.08.025

3. Rumgay H, Arnold M, Ferlay J, Lesi O, Cabasag CJ, Vignat J, Laversanne M, McGlynn KA, Soerjomataram I. Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol 2022;77(6):1598-1606. doi: 10.1016/j.jhep.2022.08.021

4. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263. doi: 10.3322/caac.21834

5. Reig M, Forner A, Rimola J, Ferrer-Fàbrega J, Burrel M, Garcia-Criado Á, Kelley RK, Galle PR, Mazzaferro V, Salem R, Sangro B, Singal AG, Vogel A, Fuster J, Ayuso C, Bruix J. BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. J Hepatol 2022;76(3):681-693. doi: 10.1016/j.jhep.2021.11.018

6. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, Zhu AX, Murad MH, Marrero JA. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 2018;67(1):358-380. doi: 10.1002/hep.29086

7. Roayaie S, Obeidat K, Sposito C, Mariani L, Bhoori S, Pellegrinelli A, Labow D, Llovet JM, Schwartz M, Mazzaferro V. Resection of hepatocellular cancer ≤2 cm: results from two Western centers. Hepatology 2013;57(4):1426-1435. doi: 10.1002/hep.25832

8. Villanueva A. Hepatocellular Carcinoma. N Engl J Med 2019;380(15):1450-1462. doi: 10.1056/NEJMra1713263

9. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 2017;67(2):93-99. doi: 10.3322/caac.21388

10. Johnson PJ, Berhane S, Kagebayashi C, Satomura S, Teng M, Reeves HL, O'Beirne J, Fox R, Skowronska A, Palmer D, Yeo W, Mo F, Lai P, Iñarrairaegui M, Chan SL, Sangro B, Miksad R, Tada T, Kumada T, Toyoda H. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. J Clin Oncol 2015;33(6):550-558. doi: 10.1200/jco.2014.57.9151

11. Xu X, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, Yang G, Yan X, Zhang YD, Liu XS. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. J Hepatol 2019;70(6):1133-1144. doi: 10.1016/j.jhep.2019.02.023

12. Wakabayashi T, Ouhmich F, Gonzalez-Cabrera C, Felli E, Saviano A, Agnus V, Savadjiev P, Baumert TF, Pessaux P, Marescaux J, Gallix B. Radiomics in hepatocellular carcinoma: a quantitative review. Hepatol Int 2019;13(5):546-559. doi: 10.1007/s12072-019-09973-0

13. Ji GW, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, Li XC, Wang XH. Radiomic Features at Contrast-enhanced CT Predict Recurrence in Early Stage Hepatocellular Carcinoma: A Multi-Institutional Study. Radiology 2020;294(3):568-579. doi: 10.1148/radiol.2020191470

14. Shi JY, Wang X, Ding GY, Dong Z, Han J, Guan Z, Ma LJ, Zheng Y, Zhang L, Yu GZ, Wang XY, Ding ZB, Ke AW, Yang H, Wang L, Ai L, Cao Y, Zhou J, Fan J, Liu

X, Gao Q. Exploring prognostic indicators in the pathological images of hepatocellular carcinoma based on deep learning. Gut 2021;70(5):951-961. doi: 10.1136/gutjnl-2020-320930

15. Kather JN, Calderaro J. Development of AI-based pathology biomarkers in gastrointestinal and liver cancer. Nature Reviews Gastroenterology & Hepatology 2020;17(10):591-592.

16. Saillard C, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, Pronier E, Laurent A, Amaddeo G, Regnault H. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. Hepatology 2020;72(6):2000-2013.

17. Xia T, Zhao B, Li B, Lei Y, Song Y, Wang Y, Tang T, Ju S. MRI‐based radiomics and deep learning in biological characteristics and prognosis of hepatocellular carcinoma: Opportunities and challenges. Journal of Magnetic Resonance Imaging 2024;59(3):767-783.

18. Wang K, Xiang Y, Yan J, Zhu Y, Chen H, Yu H, Cheng Y, Li X, Dong W, Ji Y. A deep learning model with incorporation of microvascular invasion area as a factor in predicting prognosis of hepatocellular carcinoma after R0 hepatectomy. Hepatology International 2022;16(5):1188-1198.

19. Ioannou GN, Tang W, Beste LA, Tincopa MA, Su GL, Van T, Tapper EB, Singal

AG, Zhu J, Waljee AK. Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis C cirrhosis. JAMA network open 2020;3(9):e2015626-e2015626.

20. Pawlik TM, Delman KA, Vauthey JN, Nagorney DM, Ng IOL, Ikai I, Yamaoka Y, Belghiti J, Lauwers GY, Poon RT. Tumor size predicts vascular invasion and histologic grade: implications for selection of surgical treatment for hepatocellular carcinoma. Liver Transplantation 2005;11(9):1086-1092.

21. Duvoux C, Roudot–Thoraval F, Decaens T, Pessione F, Badran H, Piardi T, Francoz C, Compagnon P, Vanlemmens C, Dumortier J. Liver transplantation for hepatocellular carcinoma: a model including α-fetoprotein improves the performance of Milan criteria. Gastroenterology 2012;143(4):986-994. e983.

22. Bai D-S, Zhang C, Chen P, Jin S-J, Jiang G-Q. The prognostic correlation of AFP level at diagnosis with pathological grade, progression, and survival of patients with hepatocellular carcinoma. Scientific reports 2017;7(1):12870.

23. Hasegawa K, Makuuchi M, Kokudo N, Izumi N, Ichida T, Kudo M, Ku Y, Sakamoto M, Nakashima O, Matsui O. Impact of histologically confirmed lymph node metastases on patient survival after surgical resection for hepatocellular carcinoma: report of a Japanese nationwide survey. Annals of surgery 2014;259(1):166-170.

24. Affo S, Yu L-X, Schwabe RF. The role of cancer-associated fibroblasts and fibrosis in liver cancer. Annual Review of Pathology: Mechanisms of Disease 2017;12(1):153-186.

25. Ramakrishna G, Rastogi A, Trehanpati N, Sen B, Khosla R, Sarin SK. From cirrhosis to hepatocellular carcinoma: new molecular insights on inflammation and cellular senescence. Liver cancer 2013;2(3-4):367-383.

26. Chen RJ, Lu MY, Wang J, Williamson DF, Rodig SJ, Lindeman NI, Mahmood F. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging 2020;41(4):757-770.

27. Xia H, Huang Q, Huang Z, Zhou Z, Zeng Y, Ma J, Fan X, Huang Y, Dong Y, Zhao H. Multimodal Deep Learning Model for Predicting Prognosis Following Radiotherapy-Based Combination Therapy in Unresectable Hepatocellular Carcinoma. Cancer Letters 2025:218122.

28. Arun N, Weston A, Bolan C, Toskich B, Lopez FG, Galeotti J. Multi-Phase Attention for Early Stage Hepatocellular Carcinoma Segmentation from Contrast Enhanced Liver MRI. 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI): IEEE, 2025; p. 1-5.

29. Maftouni M, Shen B, Law ACC, Yazdi NA, Hadavand F, Ghiasvand F, Kong Z. A

mask-guided attention deep learning model for COVID-19 diagnosis based on an integrated CT scan images database. IISE Transactions on Healthcare Systems Engineering 2023;13(2):132-149.

30. Liu G, Mitra D, Jones EF, Franc BL, Behr SC, Nguyen A, Bolouri MS, Wisner DJ, Joe BN, Esserman LJ. Mask-guided convolutional neural network for breast tumor prognostic outcome prediction on 3D DCE-MR images. Journal of Digital Imaging 2021;34(3):630-636.

31. Shi J, Sun D, Wu K, Jiang Z, Kong X, Wang W, Wu H, Zheng Y. Positional encoding-guided transformer-based multiple instance learning for histopathology whole slide images classification. Computer Methods and Programs in Biomedicine 2025;258:108491.

32. Shao Z, Bian H, Chen Y, Wang Y, Zhang J, Ji X. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems 2021;34:2136-2147.

33. Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, Vaidya AJ, Chen C, Zhuang L, Williamson DF. Artificial intelligence for multimodal data integration in oncology. Cancer cell 2022;40(10):1095-1110.

34. Lee C, Zame W, Yoon J, Van Der Schaar M. Deephit: A deep learning approach to survival analysis with competing risks. Proceedings of the AAAI conference

on artificial intelligence2018.

35. Kamran F, Wiens J. Estimating calibrated individualized survival curves with deep learning. Proceedings of the AAAI conference on artificial intelligence2021; p. 240-248.

36. Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems 2019;32.

37. Lee M, Chang Y, Oh S, Cho YY, Jung D-E, Kim HH, Nam JY, Cho H, Cho EJ, Lee J-H. Assessment of the surveillance interval at 1 year after curative treatment in hepatocellular carcinoma: risk stratification. Gut and Liver 2018;12(5):571.

38. Cucchetti A, Trevisani F, Cescon M, Ercolani G, Farinati F, Del Poggio P, Rapaccini G, Di Nolfo MA, Benvegnù L, Zoli M. Cost-effectiveness of semi-annual surveillance for hepatocellular carcinoma in cirrhotic patients of the Italian Liver Cancer population. Journal of hepatology 2012;56(5):1089-1096.

Figures

Figure 1

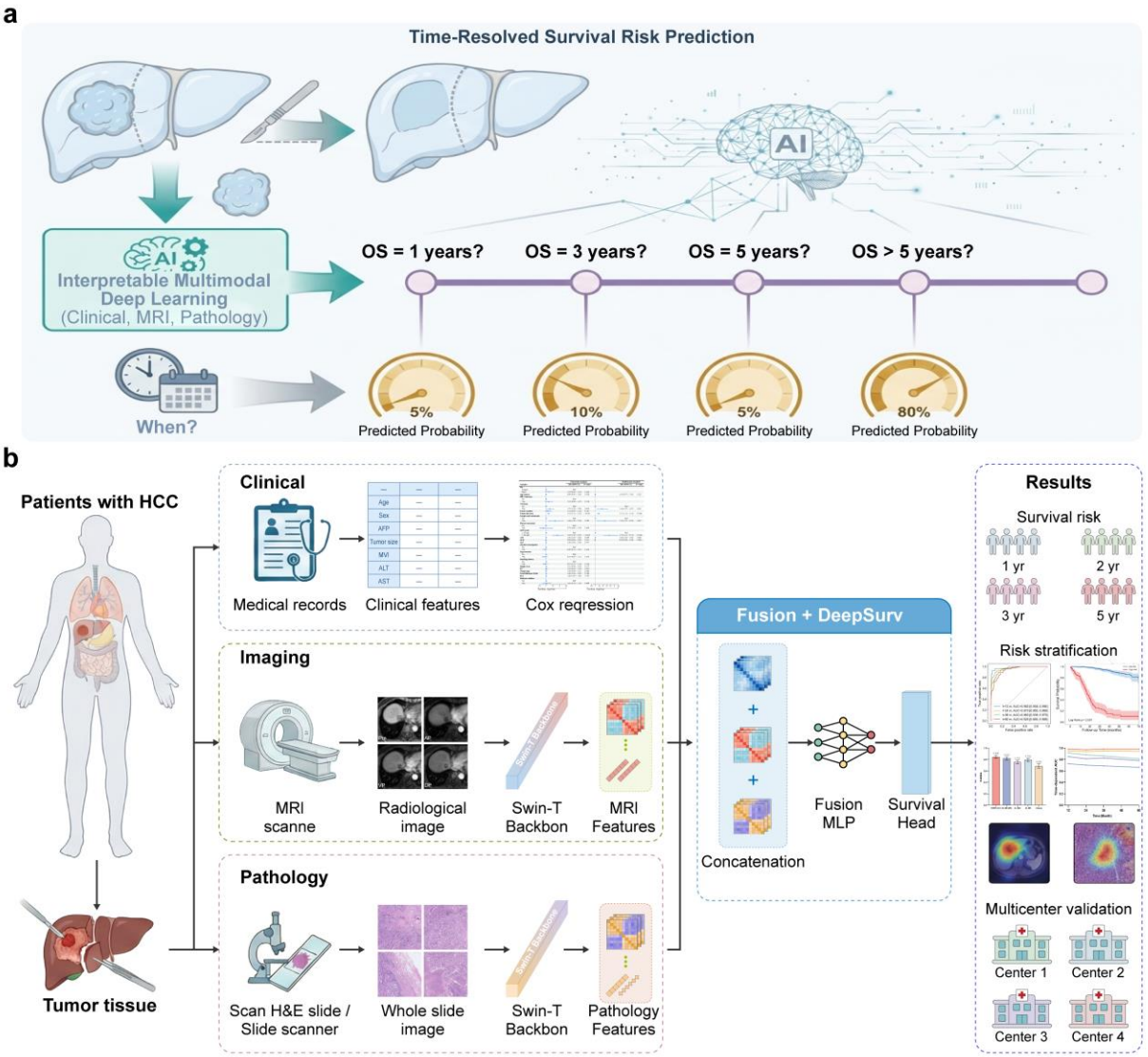

a
Time-Resolved Survival Risk Prediction
AI
Interpretable Multimodal Deep Learning (Clinical, MRI, Pathology)
When?
OS = 1 years?
OS = 3 years?
OS = 5 years?
OS > 5 years?
5% Predicted Probability
10% Predicted Probability
5% Predicted Probability
80% Predicted Probability
b
Patients with HCC
Tumor tissue
Clinical
Medical records
Clinical features
Cox reqression
Imaging
MRI scanne
Radiological image
Swin-T Backbon
MRI Features
Pathology
Scan H&E slide / Slide scanner
Whole slide image
Swin-T Backbon
Pathology Features
Fusion + DeepSurv
Concatenation
Fusion MLP
Survival Head
Results
Survival risk
1 yr
2 yr
3 yr
5 yr
Risk stratification
Multicenter validation
Center 1
Center 2
Center 3
Center 4

Figure 2

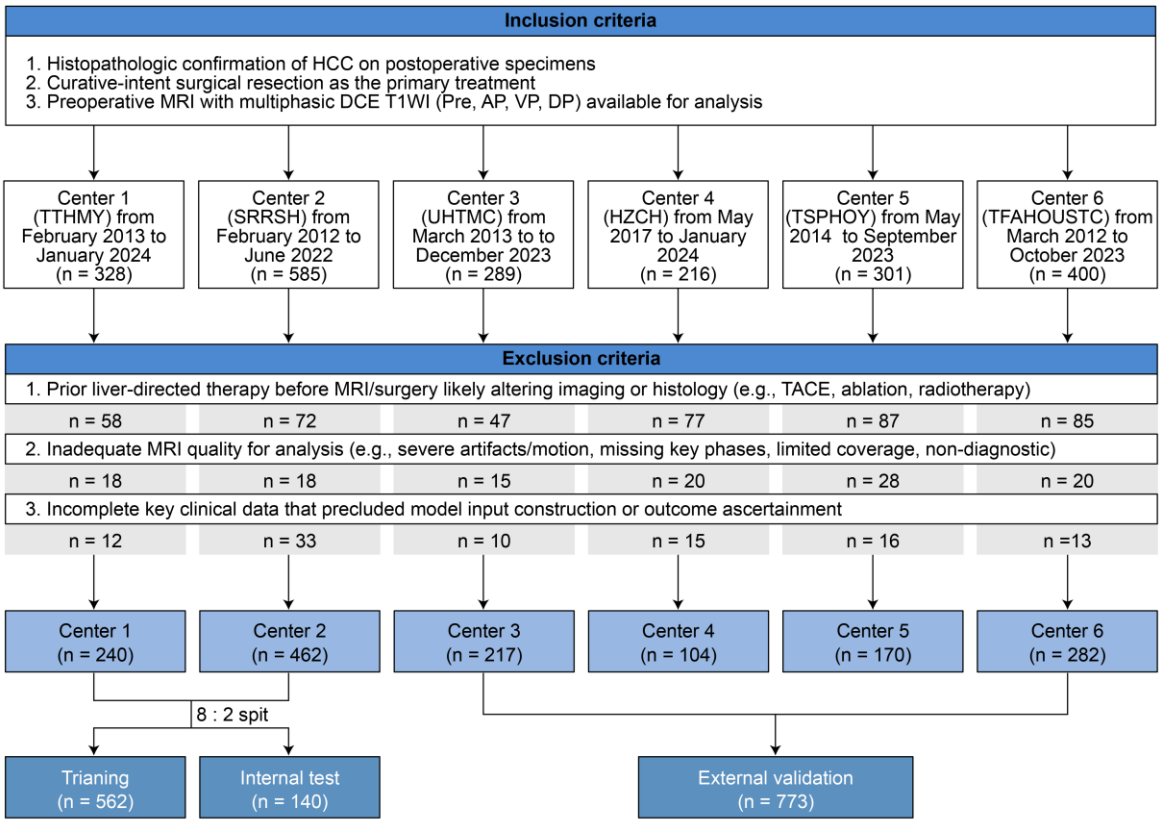
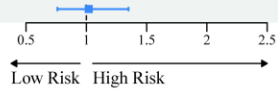
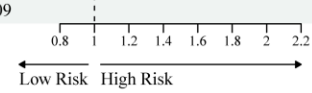


Figure 3

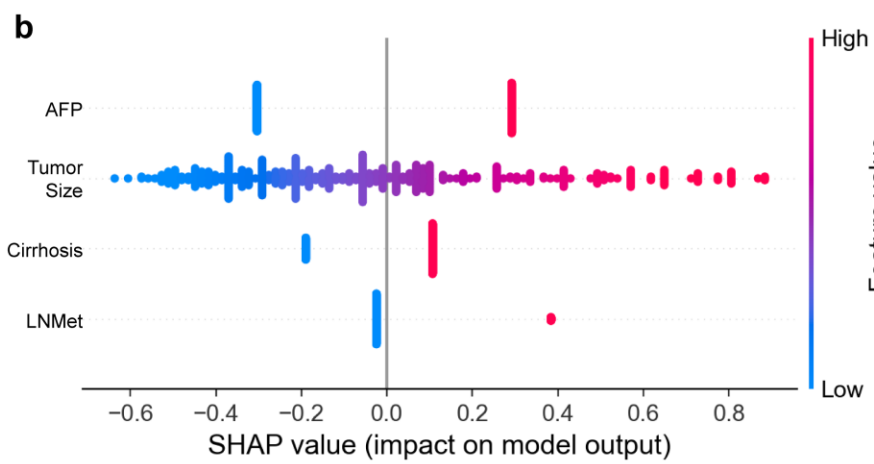
a

Variable	Univariate Analysis HR (95% CI)	*P*− value	Multivariate Analysis HR (95% CI)	*P*−value
Sex				
Female	Ref			
Male	1.19 (0.90 − 1.57)	0.234		
Age (years)	0.99 (0.98 − 1.00)	0.026	1.00 (0.97 − 1.00)	0.223
HBV Infection				
No	Ref			
Yes	1.18 (0.94 − 1.47)	0.156		
Cirrhosis				
No	Ref		Ref	
Yes	1.27 (1.02 − 1.59)	0.032	1.34 (1.07 − 1.67)	0.011
Tumor number	1.14 (0.95 − 1.37)	0.147		
Tumor size (cm)	1.18 (1.13 − 1.23)	< 0.001	1.17 (1.12 − 1.22)	< 0.001
Lymph node metastasis				
No	Ref		Ref	
Yes	1.66 (1.20 − 2.30)	0.002	1.50 (1.07 − 2.08)	0.017
Distant metastasis				
No	Ref			
Yes	0.78 (0.40 − 1.50)	0.449		
AFP Level				
≤ 20 ng/L	Ref		Ref	
> 20 ng/L	1.85 (1.50 − 2.27)	< 0.001	1.74 (1.41 − 2.15)	< 0.001
AST	1.00 (1.00 − 1.00)	0.009	1.00 (1.00 − 1.00)	0.085
ALB	0.98 (0.96 − 1.00)	0.011	0.98 (0.96 − 1.00)	0.056
ALT	1.00 (1.00 − 1.00)	0.077		
Alcohol consumption				
No	Ref			
Yes	0.89 (0.73 − 1.07)	0.252		
Hypertension				
No	Ref			
Yes	0.87 (0.68 − 1.11)	0.255		
Smoking history				
No	Ref			
Yes	0.91 (0.75 − 1.10)	0.328		
Height (cm)	1.01 (0.99 − 1.02)	0.517		
PT	1.02 (0.94 − 1.10)	0.627		
Weight (kg)	1.00 (0.99 − 1.01)	0.686		
Total Bilirubin (TBil)	1.00 (1.00 − 1.00)	0.704		
PLT	1.00 (1.00 − 1.00)	0.705		
Diabetes mellitus				
No	Ref			
Yes	1.02 (0.76 − 1.35)	0.909		





b



c

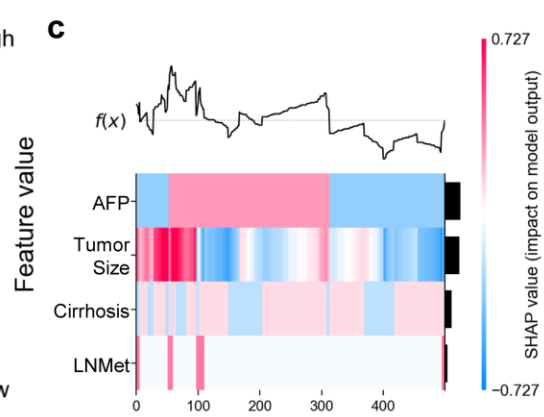


Figure 4

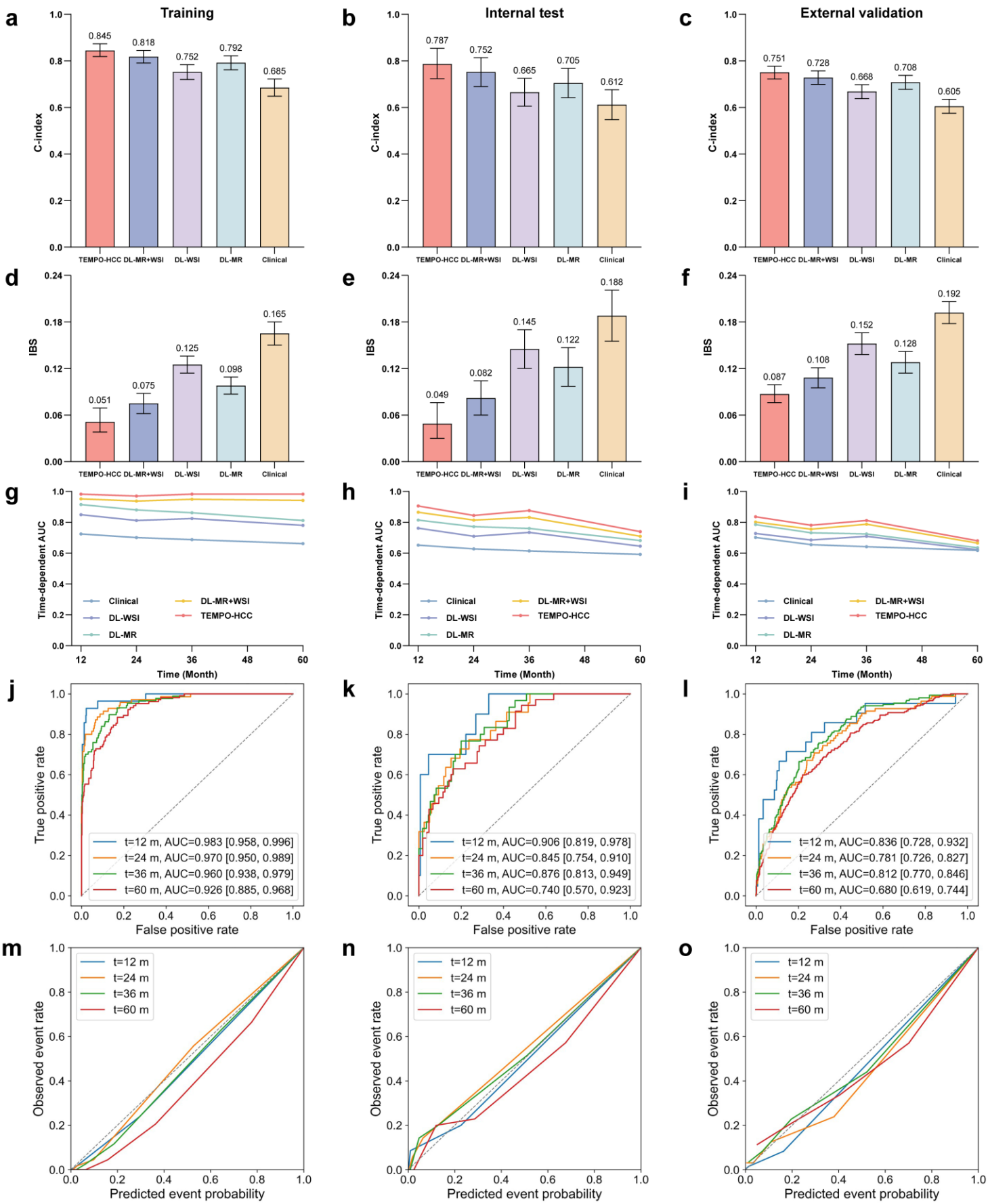

a
Training
C-index
0.845
0.818
0.752
0.792
0.685
b
Internal test
C-index
0.787
0.752
0.665
0.705
0.612
c
External validation
C-index
0.751
0.728
0.668
0.708
0.605
TEMPO-HCC
DL-MR+WSI
DL-WSI
DL-MR
Clinical
d
IBS
0.051
0.075
0.125
0.098
0.165
e
IBS
0.049
0.082
0.145
0.122
0.188
f
IBS
0.087
0.108
0.152
0.128
0.192
g
h
i
Time-dependent AUC
Time (Month)
Clinical
DL-WSI
DL-MR
DL-MR+WSI
TEMPO-HCC
j
k
l
True positive rate
False positive rate
t=12 m, AUC=0.983 [0.958, 0.996]
t=24 m, AUC=0.970 [0.950, 0.989]
t=36 m, AUC=0.960 [0.938, 0.979]
t=60 m, AUC=0.926 [0.885, 0.968]
t=12 m, AUC=0.906 [0.819, 0.978]
t=24 m, AUC=0.845 [0.754, 0.910]
t=36 m, AUC=0.876 [0.813, 0.949]
t=60 m, AUC=0.740 [0.570, 0.923]
t=12 m, AUC=0.836 [0.728, 0.932]
t=24 m, AUC=0.781 [0.726, 0.827]
t=36 m, AUC=0.812 [0.770, 0.846]
t=60 m, AUC=0.680 [0.619, 0.744]
m
n
o
Observed event rate
Predicted event probability
t=12 m
t=24 m
t=36 m
t=60 m

Figure 5

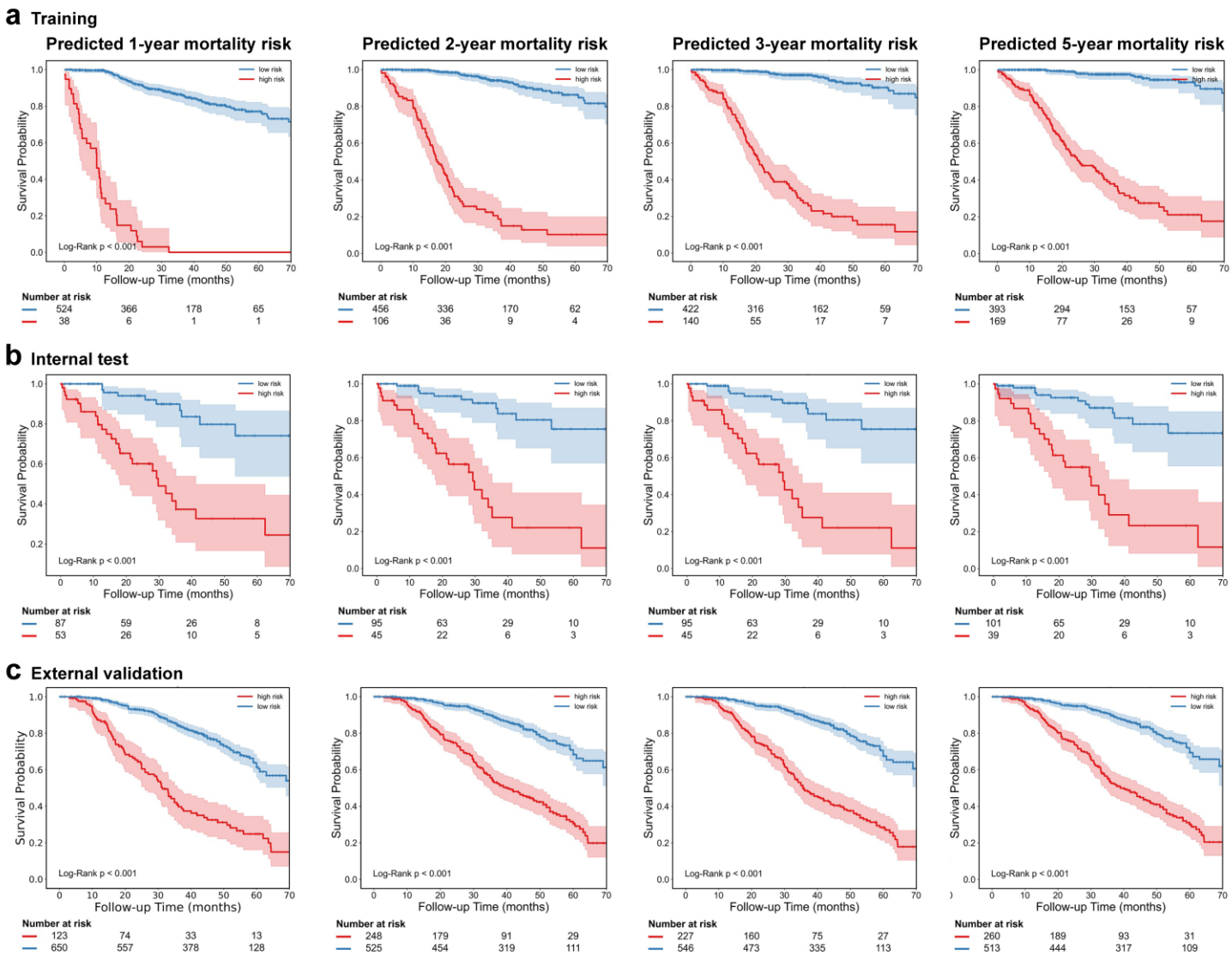

a Training
Predicted 1-year mortality risk
Predicted 2-year mortality risk
Predicted 3-year mortality risk
Predicted 5-year mortality risk
b Internal test
c External validation
Survival Probability
Follow-up Time (months)
low risk
high risk
Log-Rank p < 0.001
Number at risk
524 366 178 65
38 6 1 1
456 336 170 62
106 36 9 4
422 316 162 59
140 55 17 7
393 294 153 57
169 77 26 9
87 59 26 8
53 26 10 5
95 63 29 10
45 22 6 3
95 63 29 10
45 22 6 3
101 65 29 10
39 20 6 3
123 74 33 13
650 557 378 128
248 179 91 29
525 454 319 111
227 160 75 27
546 473 335 113
260 189 93 31
513 444 317 109

Figure 6

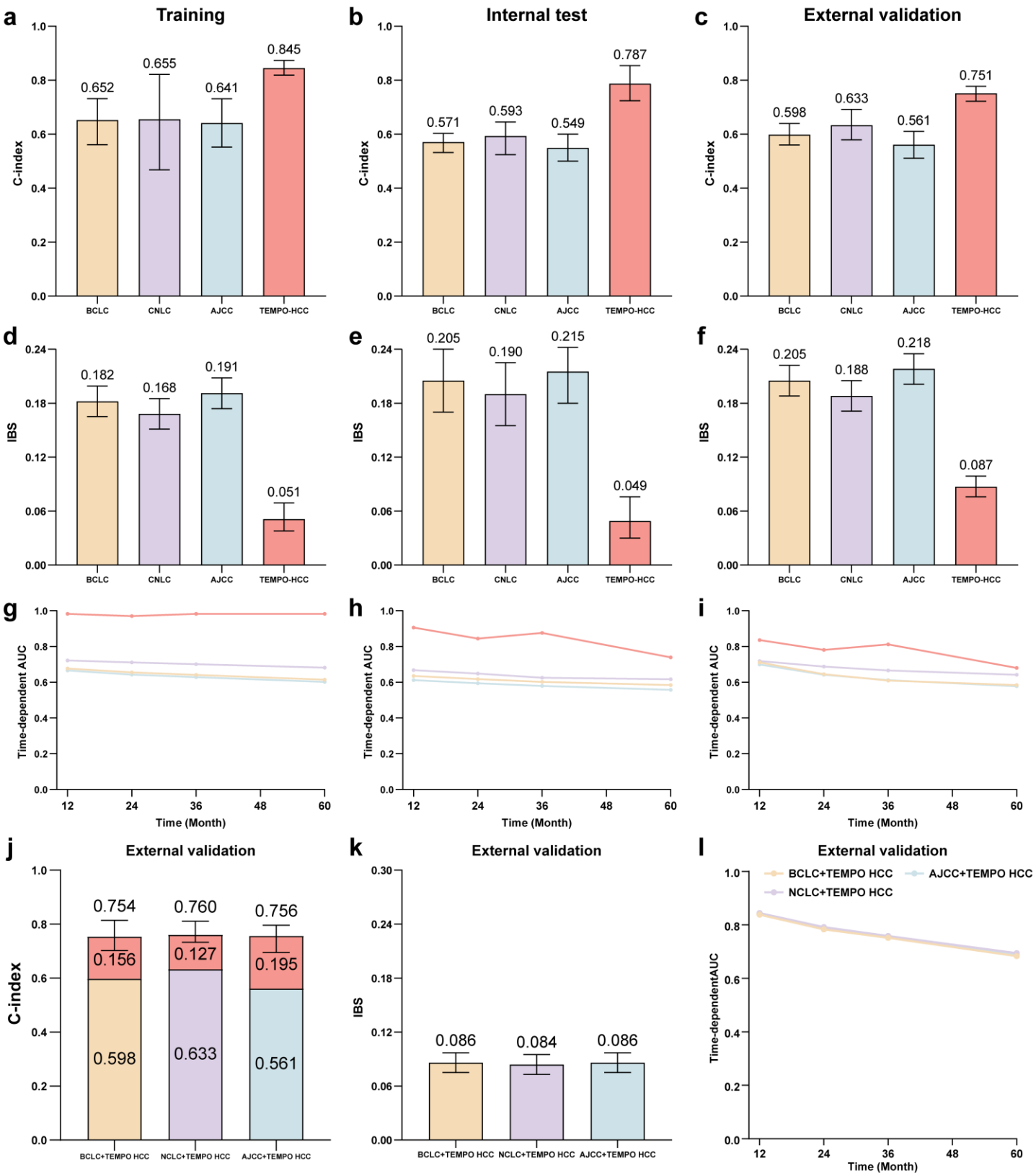
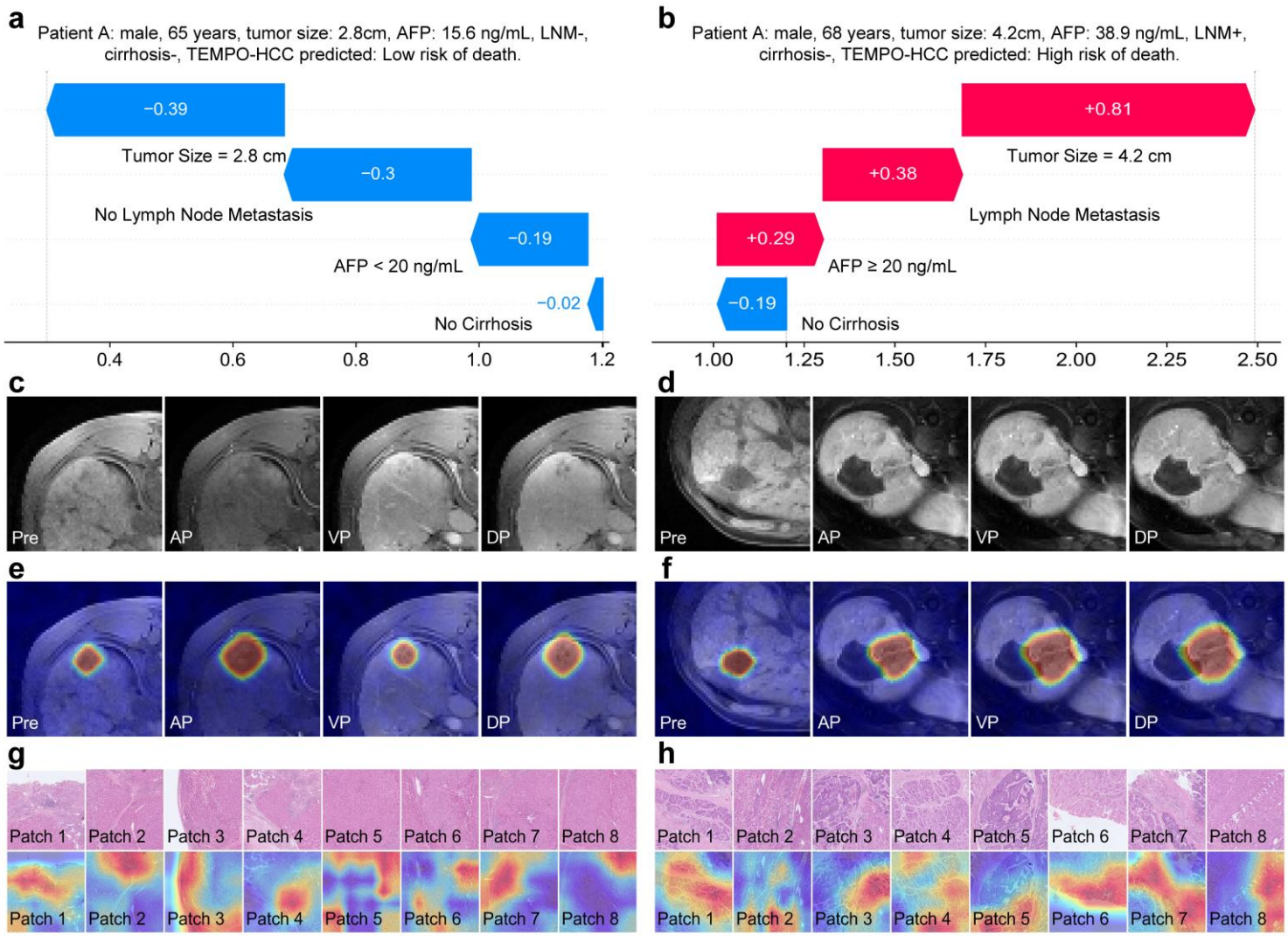

a
Training
C-index
0.652
0.655
0.641
0.845
BCLC
CNLC
AJCC
TEMPO-HCC
b
Internal test
C-index
0.571
0.593
0.549
0.787
BCLC
CNLC
AJCC
TEMPO-HCC
c
External validation
C-index
0.598
0.633
0.561
0.751
BCLC
CNLC
AJCC
TEMPO-HCC
d
IBS
0.182
0.168
0.191
0.051
BCLC
CNLC
AJCC
TEMPO-HCC
e
IBS
0.205
0.190
0.215
0.049
BCLC
CNLC
AJCC
TEMPO-HCC
f
IBS
0.205
0.188
0.218
0.087
BCLC
CNLC
AJCC
TEMPO-HCC
g
Time-dependent AUC
Time (Month)
h
Time-dependent AUC
Time (Month)
i
Time-dependent AUC
Time (Month)
j
External validation
C-index
0.754
0.760
0.756
0.156
0.127
0.195
0.598
0.633
0.561
BCLC+TEMPO HCC
NCLC+TEMPO HCC
AJCC+TEMPO HCC
k
External validation
IBS
0.086
0.084
0.086
BCLC+TEMPO HCC
NCLC+TEMPO HCC
AJCC+TEMPO HCC
l
External validation
BCLC+TEMPO HCC
AJCC+TEMPO HCC
NCLC+TEMPO HCC
Time-dependentAUC
Time (Month)

Figure 7


a
Patient A: male, 65 years, tumor size: 2.8cm, AFP: 15.6 ng/mL, LNM-, cirrhosis-, TEMPO-HCC predicted: Low risk of death.
−0.39
Tumor Size = 2.8 cm
−0.3
No Lymph Node Metastasis
−0.19
AFP < 20 ng/mL
−0.02
No Cirrhosis
0.4
0.6
0.8
1.0
1.2
b
Patient A: male, 68 years, tumor size: 4.2cm, AFP: 38.9 ng/mL, LNM+, cirrhosis-, TEMPO-HCC predicted: High risk of death.
+0.81
Tumor Size = 4.2 cm
+0.38
Lymph Node Metastasis
+0.29
AFP ≥ 20 ng/mL
−0.19
No Cirrhosis
1.00
1.25
1.50
1.75
2.00
2.25
2.50
c
d
e
f
Pre
AP
VP
DP
g
h
Patch 1
Patch 2
Patch 3
Patch 4
Patch 5
Patch 6
Patch 7
Patch 8

Tables

Table 1. Baseline demographic, clinical, laboratory, imaging, and pathologic characteristics of the study population by cohort

Variables	Training cohort (n=562)	Internal test cohort (n=140)	External validation cohort (n=773)
Gender			
Female	15.12% (85/562)	10.71% (15/140)	17.59% (136/773)
Male	84.88% (477/562)	89.29% (125/140)	82.41% (637/773)
Age (year)			
Mean±SD	59.15±11.51	60.82±10.96	58.67±10.66
Range	(25–89)	(30–87)	(25–86)
Liver Etiology			
HBV	79.54% (447/562)	76.43% (107/140)	70.12% (542/773)
Cirrhosis	70.82% (398/562)	73.57% (103/140)	70.12% (542/773)
Number of tumors			
1	87.72% (493/562)	90.00% (126/140)	75.42% (583/773)
2	9.07% (51/562)	9.29% (13/140)	19.28% (149/773)
≥3	3.20% (18/562)	0.71% (1/140)	5.30% (41/773)
Tumor Size (cm)			
Mean±SD	4.31±2.30	4.13±2.38	3.98±1.91
Range	(0.3–10.0)	(0.3–10.0)	(0.5–10.0)
Macro-VI			
Absent	92.35% (519/562)	96.43% (135/140)	87.06% (673/773)
Present	7.65% (43/562)	3.57% (5/140)	12.94% (100/773)

LN Metastasis			
Absent	93.24% (524/562)	93.57% (131/140)	92.88% (718/773)
Present	6.76% (38/562)	6.43% (9/140)	7.12% (55/773)
Distant Metastasis			
Absent	97.33% (547/562)	98.57% (138/140)	97.15% (751/773)
Present	2.67% (15/562)	1.43% (2/140)	2.85% (22/773)
AFP (> 20 ng/mL)			
No	50.18% (282/562)	47.14% (66/140)	43.98% (340/773)
Yes	49.82% (280/562)	52.86% (74/140)	56.02% (433/773)
MVI			
Negative	64.95% (365/562)	70.71% (99/140)	55.89% (432/773)
Positive	35.05% (197/562)	29.29% (41/140)	44.11% (341/773)
Grade			
Low grade	33.45% (188/562)	35.00% (49/140)	42.04% (325/773)
Moderate/High grade	66.55% (374/562)	65.00% (91/140)	57.96% (448/773)
AST			
Mean±SD	45.12±44.57	46.45±39.07	54.39±52.77
ALT			
Mean±SD	38.67±32.15	40.26±30.10	53.97±46.89
PLT			
Mean±SD	137.55±62.37	138.08±55.85	142.08±64.26
TBil			
Mean±SD	17.72±13.11	21.97±50.12	19.31±24.68
PT			
Mean±SD	13.90±1.10	14.24±1.41	13.65±1.37
ALB			

Mean±SD	38.64±4.79	38.63±6.10	38.64±4.70
Follow-up Time Range	0.2 – 115.1	0.2 – 91.5	0.2 – 83.2
Censoring rate	80.25% (451/562)	74.29% (104/140)	66.49% (514/773)
Median follow-up time	34.9 (95% CI, 33.0 – 36.6)	30.9 (95% CI, 26.6 – 35.0)	50.0 (95% CI, 48.0 – 51.0)

Note: Patient characteristics are summarized for the training, internal test, and external validation cohorts. Categorical variables are presented as percentage (number/total), and continuous variables are presented as mean ± standard deviation with ranges where applicable. In the external validation cohort, the proportions of patients with HBV infection and cirrhosis were both 70.12% (542/773). This numerical identity reflects the actual distribution of this cohort rather than a tabulation error. Tumor burden (number of tumors and tumor size), macrovascular invasion, nodal and distant metastasis status, serum alpha-fetoprotein category, microvascular invasion, and histologic grade are reported together with routine liver function–related laboratory measures. Abbreviations: HBV, hepatitis B virus; HCC, hepatocellular carcinoma; Macro-VI, macrovascular invasion; LN, lymph node; AFP, alpha-fetoprotein; MVI, microvascular invasion; AST, aspartate aminotransferase; ALT, alanine aminotransferase; PLT, platelet count; TBil, total bilirubin; PT, prothrombin time; ALB, albumin.

Table 2. Predictive performance of TEMPO-HCC for time-resolved overall survival risk prediction across cohorts

Cohort	Model	C-index	IBS	1-Year AUC (95% CI)	2-Year AUC (95% CI)	3-Year AUC (95% CI)	5-Year AUC (95% CI)
Training cohort	Clinical	0.685 (0.648-0.722)	0.165 (0.150-0.180)	0.725 (0.681-0.769)	0.701 (0.655-0.747)	0.688 (0.642-0.734)	0.662 (0.615-0.709)
	DL-WSI	0.752 (0.720-0.784)	0.125 (0.114-0.136)	0.850 (0.812-0.888)	0.812 (0.772-0.852)	0.825 (0.785-0.865)	0.780 (0.735-0.825)
	DL-MR	0.792 (0.762-0.822)	0.098 (0.087-0.109)	0.915 (0.882-0.948)	0.880 (0.845-0.915)	0.862 (0.824-0.900)	0.812 (0.770-0.854)
	DL-MR+WSI	0.818 (0.791-0.845)	0.075 (0.062-0.088)	0.952 (0.925-0.979)	0.938 (0.910-0.966)	0.950 (0.922-0.978)	0.942 (0.912-0.972)
	TEMPO-HCC	0.845 (0.819,0.873)	0.051 (0.038-0.069)	0.983 (0.958-0.996)	0.970 (0.950-0.989)	0.960 (0.938-0.979)	0.926 (0.885,0.968)
Internal test cohort	Clinical	0.612 (0.548-0.676)	0.188 (0.155-0.221)	0.652 (0.585-0.719)	0.628 (0.560-0.696)	0.615 (0.545-0.685)	0.592 (0.520-0.664)
	DL-WSI	0.665 (0.605-0.725)	0.145 (0.120-0.170)	0.762 (0.688-0.836)	0.710 (0.632-0.788)	0.735 (0.655-0.815)	0.645 (0.558-0.732)
	DL-MR	0.705 (0.642-0.768)	0.122 (0.097-0.147)	0.815 (0.745-0.885)	0.772 (0.698-0.846)	0.760 (0.685-0.835)	0.682 (0.598-0.766)
	DL-	0.752	0.082	0.865	0.815	0.832	0.710

Cohort	Model	C-index	IBS	1-Year AUC (95% CI)	2-Year AUC (95% CI)	3-Year AUC (95% CI)	5-Year AUC (95% CI)
External validation cohort	MR+WSI	(0.690-0.814)	(0.060-0.104)	(0.802-0.928)	(0.748-0.882)	(0.765-0.899)	(0.615-0.805)
	TEMPO-HCC	0.787 (0.724-0.854)	0.049 (0.030-0.076)	0.906 (0.819-0.978)	0.845 (0.754-0.910)	0.876 (0.813-0.949)	0.740 (0.570-0.923)
	Clinical	0.605 (0.575-0.635)	0.192 (0.178-0.206)	0.702 (0.655-0.749)	0.655 (0.605-0.705)	0.642 (0.590-0.694)	0.618 (0.565-0.671)
	DL-WSI	0.668 (0.638-0.698)	0.152 (0.138-0.166)	0.728 (0.682-0.774)	0.685 (0.635-0.735)	0.710 (0.662-0.758)	0.622 (0.570-0.674)
	DL-MR	0.708 (0.678-0.738)	0.128 (0.114-0.142)	0.785 (0.742-0.828)	0.732 (0.685-0.779)	0.725 (0.678-0.772)	0.635 (0.585-0.685)
	DL-MR+WSI	0.728 (0.699-0.757)	0.108 (0.095-0.121)	0.802 (0.765-0.839)	0.755 (0.712-0.798)	0.788 (0.745-0.831)	0.665 (0.618-0.712)
	TEMPO-HCC	0.751 (0.722-0.777)	0.087 (0.076-0.099)	0.836 (0.728-0.932)	0.781 (0.726-0.827)	0.812 (0.770-0.846)	0.680 (0.619-0.744)

Note: Model performance is reported for the training, internal testing, and external validation cohorts. Discrimination is summarized by Harrell's concordance index (C-index) and time-dependent area under the receiver operating characteristic curve (AUC) at 1, 2, 3, and 5 years. Overall prediction error is quantified using the integrated Brier score (IBS). For each metric, values are presented with 95% confidence intervals estimated by bootstrap resampling. Abbreviations: TEMPO-HCC, Time-resolved Explainable Multimodal Prognostic Model for Hepatocellular Carcinoma; C-index, concordance index;

AUC, area under the receiver operating characteristic curve; IBS, integrated Brier score; CI, confidence interval.

Figure legends

Figure 1. Study overview and architecture of the Time-resolved Explainable Multimodal Prognostic Model for Hepatocellular Carcinoma (TEMPO-HCC). a: Conceptual schematic of time-resolved postoperative prognostication after curative-intent resection for hepatocellular carcinoma (HCC). Rather than providing a single binary prediction, TEMPO-HCC estimates individualized overall survival risk across multiple postoperative horizons, addressing the clinical question of when adverse outcomes are most likely to occur. b: Overall workflow of TEMPO-HCC. Clinical variables, preoperative multiparametric MRI, and histopathological whole-slide images (WSIs) were integrated

into a multimodal model to generate individualized time-resolved risk predictions, which were further evaluated in internal and multicenter external validation cohorts.

Figure 2. Multicenter patient enrollment, exclusions, and cohort partitioning. Flow diagram summarizing patient screening across participating institutions, prespecified exclusions, and cohort partitioning into a derivation cohort (for training and internal test) and an independent external validation cohort for assessing generalizability.

Figure 3. Screening clinical risk factors. a: Forest plot from multivariable Cox proportional hazards regression summarizing associations between candidate perioperative variables and OS, reported as hazard ratio (HR) with 95% confidence interval (CI). b, c: Cohort-level clinical attributions using SHAP, including a patient-wise attribution heatmap (b) and a SHAP summary (beeswarm) plot (c).

Figure 4. Model performance. a–f: Bar plots depict the C-index and IBS values for the TEMPO-HCC, MRI+WSI, WSI, MRI, and Clinical models in the training, internal test, and external validation cohorts. The C-index and its associated 95% CI were calculated using the bootstrap method with 1000 resamples. In each plot, the height of the bars represents the C-index (a–c) and IBS (d–f), respectively, and the error bars represent the 95% CI. g–i: Time-dependent AUC curves showing discrimination over follow-up for TEMPO-HCC and single- and dual-modality models in overall survival (OS) prediction across cohorts. j–l: Time-dependent ROC curves of the TEMPO-HCC at multiple postoperative time points across cohorts. m–o: Calibration curves of the TEMPO-HCC model comparing

predicted event probabilities with observed event rates; the diagonal line indicates ideal calibration.

Figure 5. Risk stratification by TEMPO-HCC–predicted time-resolved mortality risk. Kaplan–Meier curves comparing high- and low-risk groups defined by predicted mortality risk at prespecified horizons. Panels show stratification based on predicted 1-, 2-, 3-, and 5-year risk in the training (a–d), internal test (e–h), and external validation (i–l) cohorts. Group differences were assessed using the log-rank test.

Figure 6. Comparison of TEMPO-HCC with three major guideline-based staging systems. a–f: Bar plots depict the C-index and IBS values for the AJCC, BCLC, CNLC, and TEMPO-HCC in the training, internal test, and external validation cohort. In each plot, the height of the bars represents the C-index (a–c) and IBS (d–f), and the error bars represent the 95% CI. g–i: Time-dependent AUC curves showing discrimination over follow-up for AJCC, BCLC, CNLC, and TEMPO-HCC models in OS prediction across cohorts. TEMPO-HCC consistently outperformed all three guideline-based staging systems. Furthermore, supplementing each of the three staging systems with TEMPO-HCC resulted in improved the C-index, IBS and AUCs in the external validation cohort (j–l). AJCC = American Joint Committee on Cancer, BCLC = Barcelona Clinic Liver Cancer, CNLC = China Liver Cancer.

Figure 7. Multi-level interpretability of TEMPO-HCC across clinical features, MRI, and WSIs. a–h: Patient-specific, cross-modality explanations for two representative

cases: panels a, c, e, and g correspond to the same low-risk HCC case, whereas panels b, d, f, and h correspond to the same high-risk HCC case. a, b: SHAP values for clinical features driving the predicted OS risk. c–f: Grad-CAM visualizations on multiphasic MRI. g, h: Representative patches from H&E-stained WSIs with corresponding saliency maps. On the basis of the importance scores, the deep red areas suggest greater contributions to the prediction, and the deep blue areas indicate lower contributions.