

Supplementary Materials

List of Supporting Information

Table S1. Participating centers

Table S2. Acquisition parameters for contrast-enhanced T1-weighted imaging across centers

Table S3. Comparisons between TEMPO-HCC and three major guideline methods

Table S4. Subgroup analyses of TEMPO-HCC in External validation cohort

Figure S1. Architecture of TEMPO-HCC for multimodal, time-resolved postoperative survival risk prediction

Figure S2. Subgroup analyses of TEMPO-HCC in External validation cohort

Figure S3. Decision curve analysis for TEMPO-HCC

Supplementary Note 1. Rationale for selecting overall survival as the sole endpoint

Supplementary Note 2. Handling of inter-institutional heterogeneity

Table S1. Participating centers

Center 1	The third hospital of mianyang (TTHMY)
Center 2	Sir Run-Run Shaw Hospital, Zhejiang University School of Medicine (SRRSH)
Center 3	Union Hospital, Tongji Medical College, Huazhong University of Science and Technology (UHTMC)
Center 4	Huzhou Central Hospital, Affiliated to Huzhou University (HZCH)
Center 5	The Second People's Hospital of YiBin (TSPHOY)
Center 6	The First Affiliated Hospital of USTC, Division of Life Science and Medicine, University of Science and Technology of China (TFAHOUSTC)

The participating centers are denoted by numeric identifiers, with corresponding abbreviations.

Table S2. Acquisition parameters for contrast-enhanced T1-weighted imaging

across centers

Center	MR-Scanner	TR (ms)	TE (ms)	FOV (mm ²)	Matrix	Section thickness (mm)	Flip angle (°)
1	Siemens 3.0T (Skyra)	3.9	1.2	420×260	320×200	3	12
1	Siemens 3.0T (Magnetom)	3.6	1.3	380×309	320×240	3	9
2	GE 3.0T (Discovery)	2.6	1.2	420×420	320×260	4	12
2	Siemens 3.0T (Skyra)	4.2	2.0	380×380	320×260	3	12
3	Siemens 3.0T (Skyra)	3.6	1.3	380×309	320×240	3	9
3	Philips 3.0T (Ingenia)	3.4	1.4	380×380	380×250	4	15
4	GE 3.0T (Discovery)	2.6	1.2	420×420	320×260	4	12
4	Siemens 3.0T (Skyra)	5.7	2.4	380×296	256×200	3	12
5	GE 3.0T (Discovery)	2.6	1.2	420×420	320×260	4	12
5	United Imaging 3.0T (uMR790)	2.7	1.3	380×380	432×432	3	8
6	Siemens 3.0T (Magnetom)	3.6	1.3	380×309	320×240	3	9
6	Siemens 3.0T (Skyra)	3.6	1.3	380×309	320×240	3	9

Acquisition parameters for contrast-enhanced T1-weighted imaging across centers.

Table S3. Comparisons between TEMPO-HCC and three major guideline

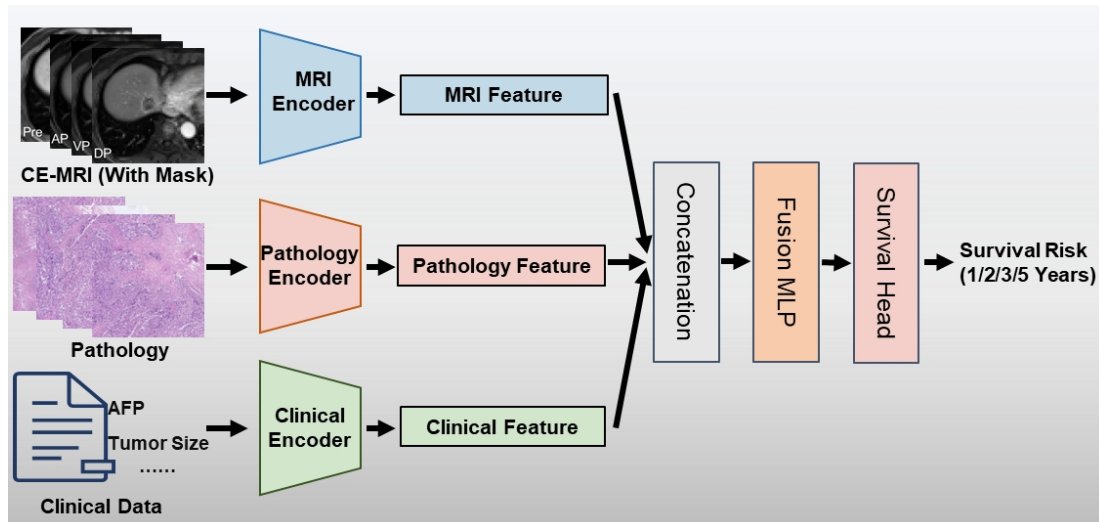
methods

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	BCLC	0.652	0.182	0.677	0.655	0.641	0.615
		(0.561-0.732)	(0.165-0.199)	(0.635-0.719)	(0.610-0.700)	(0.596-0.686)	(0.565-0.665)
	CNLC	0.655	0.168	0.722	0.711	0.701	0.682
		(0.468-0.821)	(0.151-0.185)	(0.685-0.759)	(0.672-0.750)	(0.657-0.745)	(0.632-0.732)
	AJCC	0.641	0.191	0.666	0.643	0.629	0.602
		(0.552-0.731)	(0.174-0.208)	(0.625-0.707)	(0.598-0.688)	(0.582-0.676)	(0.551-0.653)
	TEMPO-HCC	0.845	0.051	0.983	0.970	0.960	0.926
		(0.819,0.873)	(0.038-0.069)	(0.958-0.996)	(0.950-0.989)	(0.938-0.979)	(0.885,0.968)
Internal test	BCLC	0.571	0.205	0.635	0.618	0.602	0.585
		(0.532-0.603)	(0.170-0.240)	(0.571-0.699)	(0.554-0.682)	(0.541-0.663)	(0.515-0.655)
	CNLC	0.593	0.190	0.668	0.649	0.625	0.617
		(0.524-0.645)	(0.155-0.225)	(0.605-0.731)	(0.585-0.713)	(0.562-0.688)	(0.545-0.689)
	AJCC	0.549	0.215	0.612	0.595	0.580	0.558
		(0.500-0.600)	(0.180-0.242)	(0.548-0.676)	(0.530-0.660)	(0.518-0.642)	(0.495-0.621)
	TEMPO-HCC	0.787	0.049	0.906	0.845	0.876	0.740
		(0.724-0.854)	(0.030-0.076)	(0.819-0.978)	(0.754-0.910)	(0.813-0.949)	(0.570-0.923)
External validation	BCLC	0.598	0.205	0.711	0.645	0.610	0.585
		(0.560-0.640)	(0.188-0.222)	(0.661-0.761)	(0.595-0.695)	(0.566-0.654)	(0.535-0.635)
	CNLC	0.633	0.188	0.719	0.688	0.666	0.642
		(0.579-0.692)	(0.171-0.205)	(0.674-0.764)	(0.632-0.744)	(0.610-0.722)	(0.592-0.692)
	AJCC	0.561	0.218	0.700	0.642	0.612	0.578
		(0.511-0.610)	(0.201-0.235)	(0.652-0.748)	(0.590-0.694)	(0.559-0.664)	(0.525-0.631)
	TEMPO-HCC	0.751	0.087	0.836	0.781	0.812	0.680
		(0.722, 0.777)	(0.076-0.099)	(0.728-0.932)	(0.726-0.827)	(0.770-0.846)	(0.619-0.744)
External validation	BCLC+TEMPO HCC	0.754	0.086	0.839	0.784	0.752	0.684
		(0.725-0.783)	(0.075-0.097)	(0.731-0.935)	(0.729-0.830)	(0.701-0.803)	(0.623-0.748)
	CNLC+TEMPO HCC	0.760	0.084	0.845	0.792	0.759	0.695
		(0.731-0.789)	(0.073-0.095)	(0.737-0.941)	(0.737-0.838)	(0.705-0.813)	(0.634-0.759)
	AJCC+TEMPO HCC	0.756	0.086	0.841	0.787	0.755	0.688
		(0.727-0.785)	(0.075-0.097)	(0.733-0.937)	(0.732-0.833)	(0.702-0.808)	(0.627-0.752)

Table S4. Subgroup analyses of TEMPO-HCC in External validation cohort

Cohort	C-index	IBS	1-Year AUC (95% CI)	2-Year AUC (95% CI)	3-Year AUC (95% CI)	5-Year AUC (95% CI)
Center 3	0.755	0.086	0.840	0.785	0.816	0.685
	(0.710-0.800)	(0.072-0.100)	(0.742-0.938)	(0.728-0.842)	(0.772-0.860)	(0.621-0.749)
Center 4	0.712	0.098	0.788	0.745	0.772	0.615
	(0.635-0.789)	(0.078-0.118)	(0.675-0.901)	(0.670-0.820)	(0.715-0.829)	(0.535-0.695)
Center 5	0.745	0.089	0.832	0.775	0.805	0.672
	(0.698-0.792)	(0.075-0.103)	(0.725-0.939)	(0.712-0.838)	(0.762-0.848)	(0.605-0.739)
Center 6	0.765	0.083	0.850	0.795	0.825	0.702
	(0.730-0.800)	(0.072-0.094)	(0.748-0.952)	(0.738-0.852)	(0.785-0.865)	(0.640-0.764)
Cirrhosis (+)	0.742	0.089	0.825	0.772	0.804	0.665
	(0.715-0.769)	(0.080-0.098)	(0.731-0.919)	(0.724-0.820)	(0.765-0.843)	(0.612-0.718)
Cirrhosis (–)	0.772	0.082	0.862	0.802	0.832	0.715
	(0.732-0.812)	(0.070-0.094)	(0.755-0.969)	(0.742-0.862)	(0.788-0.876)	(0.645-0.785)
AFP (> 20 ng/mL)	0.780	0.080	0.871	0.812	0.842	0.725
	(0.748-0.812)	(0.069-0.091)	(0.762-0.980)	(0.755-0.869)	(0.798-0.886)	(0.660-0.790)
AFP (≤ 20 ng/mL)	0.728	0.093	0.810	0.758	0.788	0.645
	(0.699-0.757)	(0.083-0.103)	(0.718-0.902)	(0.710-0.806)	(0.748-0.828)	(0.590-0.700)
Size (≤ 3 cm)	0.788	0.078	0.882	0.821	0.852	0.738
	(0.755-0.821)	(0.066-0.090)	(0.775-0.989)	(0.765-0.877)	(0.805-0.899)	(0.675-0.801)
Size (> 3 cm)	0.729	0.094	0.808	0.755	0.785	0.643
	(0.702-0.756)	(0.085-0.103)	(0.712-0.904)	(0.705-0.805)	(0.745-0.825)	(0.588-0.698)
TEMPO-HCC	0.751	0.087	0.836	0.781	0.812	0.680
	(0.722, 0.777)	(0.076-0.099)	(0.728-0.932)	(0.726-0.827)	(0.770-0.846)	(0.619-0.744)

Figure S1. Architecture of TEMPO-HCC for multimodal, time-resolved postoperative survival risk prediction



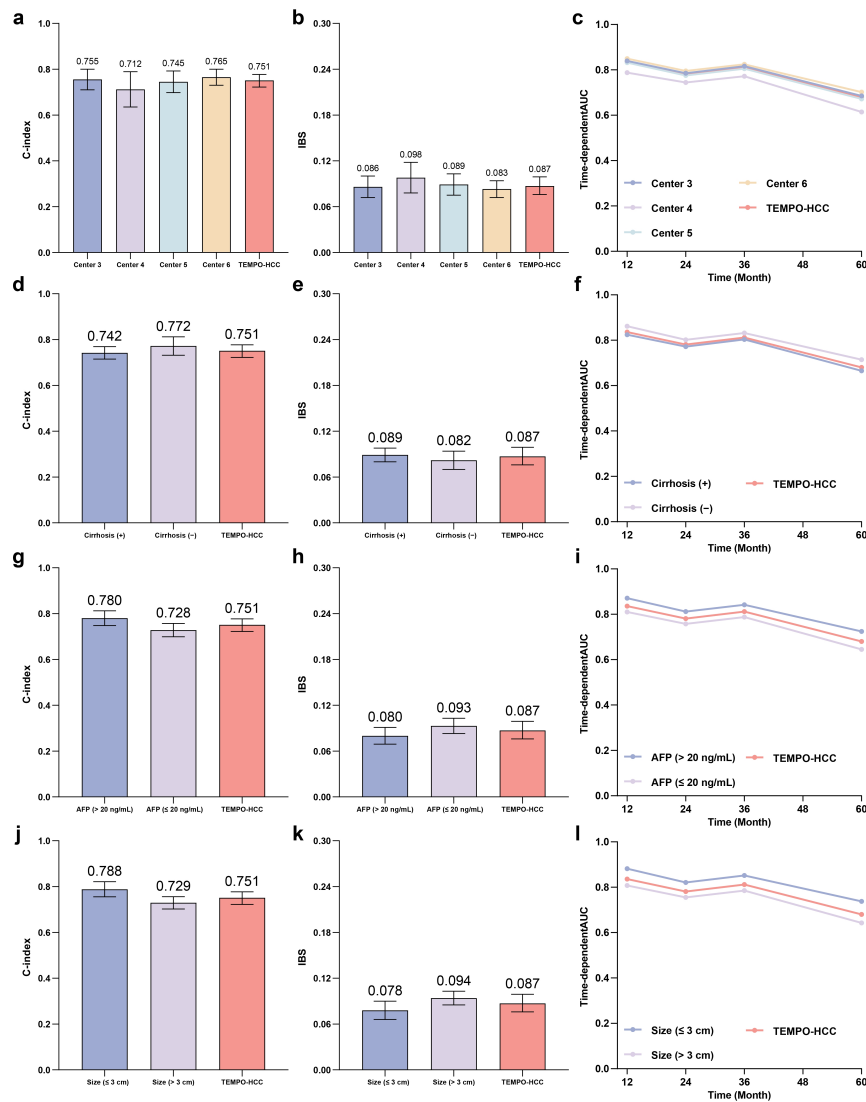
Overall network schematic showing modality-specific encoders for preoperative contrast-enhanced MRI (CE-MRI) with tumor masks, postoperative H&E histopathology, and perioperative clinical data; modality representations are fused to generate time-resolved survival risk estimates at prespecified horizons. MRI encoder with mask-guided hierarchical attention:

A Swin-Transformer Tiny (Swin-T) backbone pretrained on ImageNet was used to encode each 2D slice into a feature vector. Features were projected to a common embedding dimension (512). To incorporate tumor localization without explicit segmentation supervision, we computed a tumor fraction per slice as the mean value of the resized tumor mask. This scalar was concatenated to the slice embedding and used as input to a slice-attention network that produced normalized attention weights across the S slices within each phase. Phase-level attention subsequently aggregated the four phase

embeddings into a single MRI representation. The encoder outputs: A patient-level MRI embedding: $(B \times 512)$. Pathology encoder: Swin-T + position injection + MIL attention: Each patch was encoded by a (frozen) Swin-T backbone to obtain patch-level features. A small MLP mapped normalized patch coordinates to a positional embedding of the same dimension and was injected into patch features with a small scaling factor (0.1) to stabilize training. A MIL attention module produced attention weights over patches and aggregated them into a patient-level pathology embedding, which was further refined by an MLP projection head to 512 dimensions. Outputs: Pathology embedding: $(B \times 512)$. Clinical risk (shape $(B \times 1)$) was mapped to a low-dimensional embedding (32 dimensions) using a non-negative linear layer implemented by applying a softplus transform to raw weights. This enforces monotonicity between risk score and embedding along learned directions. Multimodal fusion and survival heads: MRI (512D), pathology (512D), and clinical embeddings (32D) were concatenated and passed through a fusion MLP with ReLU activations and dropout. A discrete-time survival head then predicted hazard logits over K time bins. Discrete-Time Survival Formulation: Time discretization: Survival time was discretized using fixed bin edges 0 1 2 2 4 3 6 6 0 months, producing $K = 4$ bins. For a patient, the model outputs hazard logits $\mathbf{z} \in \mathbb{R}^K$, with hazards $h_k = \sigma(z_k)$. Likelihood with right censoring: We optimized the negative log-likelihood of the discrete hazard model: For an event occurring in bin k : $\log L = \sum_{j < k} \log(1 - h_j) + \log(h_k)$, For censoring in

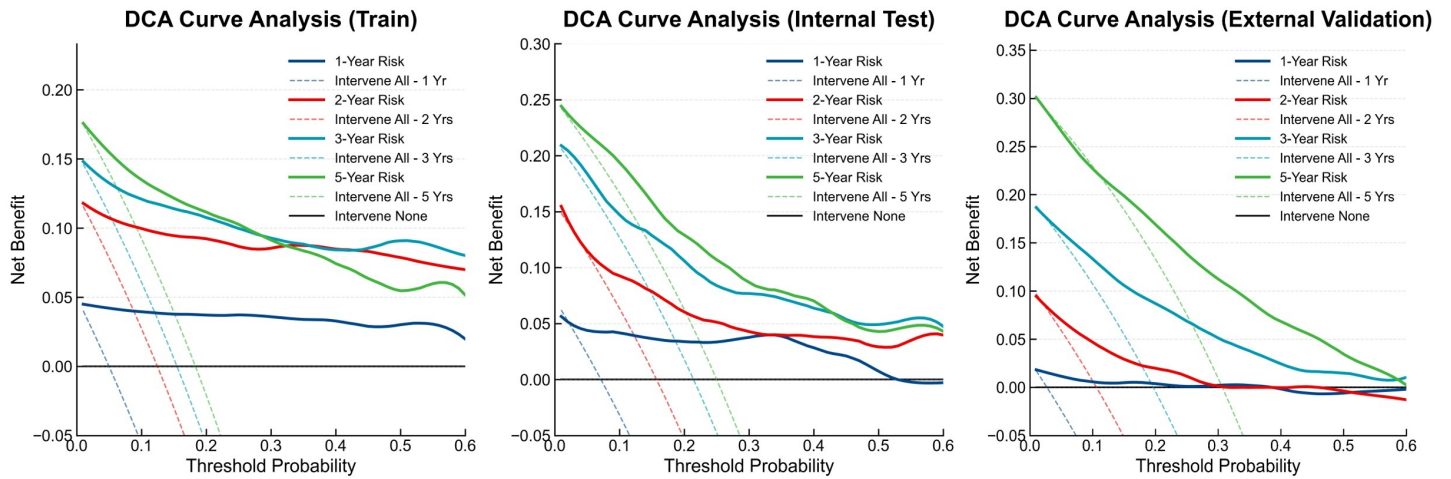
bin k : , $\log L = \sum_{j \leq k} \log(1 - h_j)$, This loss is fully differentiable, naturally supports censoring, and directly matches the model output. Risk reporting at fixed horizons: From hazards we derive the survival function $S_k = \prod_{j \leq k} (1 - h_j)$ and the cumulative incidence (risk) at horizon t by summing the probability mass of bins whose midpoints are $\leq t$. Reported 1/2/3/5-year risks correspond to 12/24/36/60 months. Optimization and Training Protocol: Optimizer: AdamW, learning rate 1×10^{-4} , weight decay 1×10^{-4} . Batch size: 2 (with variable patch padding handled by a custom collate function). Epochs: 100. Mixed precision: enabled on CUDA via automatic mixed precision and gradient scaling.

Figure S2. Subgroup analyses of TEMPO-HCC in External validation cohort



a, b, c: Discrimination performance of TEMPO-HCC in four external validation cohorts, assessed by the C-index, IBS and time-dependent AUC. d, e, f: C-index, IBS and time-dependent AUC of TEMPO-HCC in the external validation cohorts stratified by Cirrhosis status. g, h, i: C-index, IBS and time-dependent AUC of TEMPO-HCC in the external validation cohorts stratified by AFP level. j, k, l: C-index, IBS and time-dependent AUC of TEMPO-HCC in the external validation cohorts stratified by tumor size.

Figure S3. Decision curve analysis of TEMPO-HCC in the training, internal test, and external validation cohorts.



Net benefit curves of TEMPO-HCC for 1-, 2-, 3-, and 5-year risk prediction in the training cohort (left), internal test cohort (middle), and external validation cohort (right). Dashed lines indicate the corresponding treat-all strategies, and the horizontal black line indicates the treat-none strategy. Greater net benefit indicates superior potential clinical utility.

Supplementary Note 1. Rationale for selecting overall survival as the sole endpoint

Overall survival (OS) was selected as the sole primary endpoint in the present study after careful clinical and methodological consideration. First, in a multicenter retrospective cohort spanning six institutions and a relatively long study period, OS represents the most objective, consistently recorded, and cross-center comparable endpoint. In our study, OS was defined as the time from curative resection to death from any cause, and survival status was primarily ascertained through routine follow-up systems, medical record review, and, when necessary, telephone follow-up. This made OS a particularly robust endpoint for data harmonization and model development across centers.

Second, although recurrence-related outcomes such as recurrence-free survival (RFS) are highly relevant in hepatocellular carcinoma (HCC) and may better reflect postoperative tumor biology, these endpoints are more vulnerable to heterogeneity in retrospective multicenter settings. In particular, recurrence-related endpoints may be influenced by differences in surveillance intensity, imaging follow-up intervals, criteria for distinguishing residual disease from recurrence, assessment after locoregional treatment, and the timing at which first recurrence is formally confirmed. Such variability may introduce substantial noise into endpoint definition and compromise the stability of model training and external validation. By contrast, OS is a harder

and more robust endpoint, making it better suited to the current stage of multicenter model development.

Third, compared with disease-specific survival (DSS), OS avoids the additional uncertainty associated with cause-of-death attribution. In real-world retrospective cohorts, the classification of death as tumor-related or non-tumor-related may be affected by incomplete documentation, differences in adjudication practice across centers, and the frequent coexistence of competing clinical events such as liver failure, infection, or other non-cancer causes of death. As a result, DSS may be more prone to endpoint misclassification. OS, by not requiring additional attribution of the cause of death, reduces this source of bias and provides a more uniform basis for modeling.

Fourth, although RFS and DSS provide important biologic and disease-specific information, OS remains one of the most clinically meaningful and interpretable outcomes in oncology. For patients with HCC after curative resection, postoperative OS risk stratification can directly support overall prognostic assessment, follow-up resource allocation, and intensified management of high-risk individuals. This is also aligned with the intended role of TEMPO-HCC as a postoperative, time-resolved survival prediction model.

Fifth, from a methodological perspective, the present study was designed as a proof-of-concept investigation to test whether a multimodal deep learning framework integrating preoperative MRI, postoperative histopathology, and clinical variables and accommodating missing modalities could enable time-resolved survival prediction after curative resection for HCC. At this stage, focusing on a single, well-defined, and robust endpoint allows a clearer evaluation of the feasibility and value of the proposed technical framework. Simultaneous incorporation of multiple endpoints, such as OS, RFS, early recurrence, and DSS, would substantially increase modeling complexity and could further complicate interpretation because of differences in endpoint quality, ascertainment, and event definition.

Finally, the selection of OS as the sole endpoint should not be interpreted as diminishing the importance of recurrence-related outcomes in HCC. On the contrary, we fully acknowledge that future extension of this framework to RFS, early recurrence, DSS, and other biologically informative endpoints may further enhance its clinical relevance, particularly for postoperative surveillance and disease monitoring. Accordingly, recurrence-related outcomes should be regarded as an important direction for future investigation and model refinement.

Supplementary Note 2. Handling of inter-institutional heterogeneity

Given the multicenter design and the long inclusion period (2012–2024), substantial heterogeneity in MRI acquisition protocols, histopathological slide preparation, and perioperative management across institutions was expected. Rather than assuming cross-center homogeneity, we sought to reduce the influence of non-biological variation through unified preprocessing, standardized input construction, quality-control procedures, and independent multicenter external validation. For MRI, images and corresponding tumor masks were resampled to a consistent voxel spacing, followed by intensity normalization using a zero-mean, unit-variance scheme. Model inputs were then constructed using a fixed number of slices and a unified input size to reduce variability related to scanner vendors, acquisition parameters, and spatial resolution. Major MRI acquisition parameters from each center are provided in the Supplementary Materials to improve transparency. For histopathology, postoperative H&E whole-slide images were processed using a standardized patch-based workflow, including patch extraction, preprocessing, and quality control, in an effort to minimize center-related differences in slide scanning, staining quality, and image resolution. Tumor annotation across centers was performed according to a unified protocol with quality control, and difficult cases were further reviewed to improve cross-center consistency in lesion delineation.

At the model level, pretrained backbones were frozen and only lightweight fusion and prediction modules were trained, thereby reducing the number of trainable parameters and limiting overfitting to center-specific patterns. Importantly, model generalizability was assessed using independent external validation cohorts from multiple institutions, which provided a direct evaluation of robustness under real-world inter-institutional heterogeneity. We did not apply explicit cross-center harmonization or domain-adaptation techniques, such as ComBat or adversarial domain adaptation. This design choice was intentional, as we aimed to evaluate TEMPO-HCC under realistic heterogeneous conditions that more closely resemble future clinical deployment, rather than under an idealized setting created by strong post hoc harmonization. In this context, inter-institutional heterogeneity should be viewed not only as a methodological challenge, but also as a representative test environment for real-world generalizability.