Advances in Differential Equations and Control Processes 2026, 33(1), 3902.
doi: https://doi.org/10.59400/adecp3902

Article

A multimodal deep learning-based dynamic prediction model for colorectal cancer liver metastasis

Haitao Zheng[1,2], Dehui Wen[1,*], Liwei Zhang[1], Haiyong Lu[1], Xiaoyu Li[1], Yongxin Li[3]

[1] Department of Ultrasound Medicine, First Affiliated Hospital of Hebei North University, Shijiazhuang 075000, China
[2] First Clinical Medical College, Hebei North University, Shijiazhuang 075000, China
[3] Department of Thoracic Surgery, China Aerospace Science and Industry Corporation 731 Hospital, Beijing 100074, China
*** Corresponding authors:** Dehui Wen, 15530396808@163.com

CITATION

Zheng H, Wen D, Zhang L, et al. A multimodal deep learning-based dynamic prediction model for colorectal cancer liver metastasis. Advances in Differential Equations and Control Processes. 2026; 33(1): 3902. https://doi.org/10.59400/adecp3902

ARTICLE INFO

Received: 16 January 2026
Revised: 26 February 2026
Accepted: 2 March 2026
Available online: 18 March 2026

Abstract: Colorectal cancer liver metastasis (CRLM) remains a major determinant of long-term outcomes. Existing clinical models are typically static and single-modality, limiting early warning and individualized follow-up. We prospectively enrolled 300 treatment-naïve colorectal cancer patients. We collected preoperative three-phase contrast-enhanced ultrasound (CEUS) dynamic sequences, longitudinal serum marker measurements (EZH2/CD10) from preoperation through 12 months, and 35 clinical–pathological variables. The proposed Dynamic Modality Alignment Network (DMA-Net) includes (i) an imaging encoder based on an enhanced 3D-ResNet18 to extract perfusion kinetics, (ii) a molecular encoder using BiLSTM with temporal attention to model serial biomarkers, and (iii) a clinical encoder (MLP) for structured variables. A dynamic alignment module and cross-modal attention fuse modalities, followed by a discrete-time survival head that outputs month-specific conditional hazards and cumulative risks. On the held-out test set, the tri-modal model achieved an area under the curve (AUC) of 0.918 at 12 months with favorable calibration (Brier score 0.123), outperforming a traditional Cox model built from clinical variables (AUC 0.782, Brier score 0.177). Time-dependent evaluation showed stable AUCs from 3 to 12 months (0.904–0.919). Ablation experiments indicated that imaging and molecular branches contributed most to discrimination, whereas clinical variables improved calibration. Multimodal dynamic modeling integrating CEUS perfusion, longitudinal biomarkers, and clinical variables improves early warning and risk stratification for CRLM, and provides a practical framework to support personalized surveillance.

Keywords: colorectal carcinoma; hepatic metastasis; contrast-enhanced ultrasound; multimodal fusion learning; discrete-time survival analysis; dynamic risk stratification

1. Introduction

Colorectal cancer (CRC) ranks as the third most prevalent malignant tumor globally, with colorectal cancer liver metastases (CRLM) being the primary cause of patient mortality. Approximately 50% of patients develop liver metastases during disease progression [1]. Current clinically relied-upon prediction models (e.g., Fong Clinical Risk Score, MSKCC model) are predominantly based on static clinical and pathological features. They struggle to capture the spatiotemporal heterogeneity of tumors and the dynamic evolution of the tumor microenvironment, resulting in limited predictive accuracy. The area under the curve (AUC) typically ranges only between

0.65 and 0.75 [2]. The explosive growth of multimodal data has driven an urgent need for deep learning integration and dynamic modeling.

This study aims to develop a multimodal deep learning model for the dynamic prediction of CRLM. Theoretically, the model overcomes the limitations of single-modality and static features [3] by integrating preoperative and follow-up dynamic contrast-enhanced ultrasound perfusion images, molecular time-series data (e.g., serum EZH2/CD10), and key clinical characteristics (e.g., TNM staging, EMVI, CEA). Through a unified deep learning framework, it characterizes the temporal evolution of liver metastasis risk, enabling personalized dynamic prediction across different postoperative time windows (e.g., 4, 6, 12 months). Clinically, existing risk models struggle to meet personalized treatment demands—for instance, significant variations in response to identical neoadjuvant chemotherapy and targeted drug regimens exist across patients [4]. The proposed dynamic risk stratification (high-risk, progression, and remission state discrimination) holds promise for optimizing neoadjuvant chemotherapy cycles, surgical timing, and postoperative follow-up strategies, thereby improving long-term outcomes for CRLM patients [5]. Technically, this study introduces dynamic time series regularization and cross-modal attention mechanisms based on multimodal temporal data.

Currently, the development of single-modality prediction models has encountered significant bottlenecks. International research has largely focused on the field of radiomics. For instance, teams such as MSKCC have developed prediction models based on CT texture and wavelet features, achieving an AUC of approximately 0.78, but these models have yet to effectively integrate molecular-level data [6, 7]. EU collaborative research has constructed predictive models using DCE-MRI kinetic parameters (AUC ~0.81), yet sensitivity toward micrometastases remains insufficient [8]. Domestic scholars have also explored PET/CT and CT radiomics-based approaches, but these generally suffer from reliance on expensive equipment and lack of dynamic validation [9]. Multimodal fusion is emerging as a new trend. At the feature level, the Pathomic Fusion framework combines H&E histopathology images with genomic data for cancer diagnosis and prognosis assessment [3, 10]. At the decision level, Transformer-based frameworks have demonstrated promising performance in multimodal medical tasks, though they remain constrained by high computational costs [11]. Concurrently, the general medical AI paradigm is advancing rapidly. Nevertheless, challenges such as multimodal data alignment, modeling temporal dependencies, and high annotation costs remain significant bottlenecks [12].

Regarding dynamic predictive models, preliminary explorations have been initiated. At the level of time series analysis, biomarkers that change over time (such as CEA) have been demonstrated to enhance the efficacy of risk assessment for recurrence and metastasis [13]. At the level of spatiotemporal joint modeling, graph learning and self-supervised learning methods have shown potential in modeling pathological progression [14,15], but systematic research on quantifying the "tumor-immune" spatial interaction mechanisms and their closed-loop optimization remains lacking. Domestically, positive outcomes have been achieved in fields

like radiomics and liquid biopsy [16, 17]. However, significant challenges persist, including pronounced data silos across modalities and insufficient prospective clinical validation [18].

Despite progress in multimodal learning and dynamic prediction research, four major challenges persist: First, insufficient multimodal dynamic coordination capabilities. Current fusion strategies emphasize static integration and lack specialized modules for temporal heterogeneity, leading to significant model performance degradation when modalities are missing [19]. Second, spatio-temporal evolution modeling remains rudimentary. Existing graph models often rely on predefined topologies, struggling to accurately capture complex spatial interactions such as vascular invasion and sinusoidal penetration. This necessitates introducing graph structure generation mechanisms based on biophysical constraints [20]. Third, there exists a disconnect between interpretability and clinical application. General interpretability techniques often fail to align with clinical treatment decision pathways, necessitating the establishment of a trinity causal attribution framework integrating "imaging-pathology-genetics" [21]. Fourth, the validation system remains incomplete. International models are predominantly based on single-center or Western public cohorts, lacking validation efficacy for Chinese populations, while domestic prospective studies still face limitations in sample size and scenario coverage [22].

The core innovation of this project lies in proposing the Dynamic Modality Alignment Network (DMA-Net). This network employs the Differentiable Dynamic Time Warping (DDTW) method to align multi-source temporal data, integrating cross-modal attention mechanisms for end-to-end fusion of imaging, molecular, and clinical features. This study prospectively enrolled 300 treatment-naive colorectal cancer patients in a single-center cohort. We systematically collected preoperative three-phase CEUS dynamic videos, preoperative and postoperative EZH2/CD10 time-series data, and 35 clinical features to construct a perioperative multimodal dynamic database. Research objectives were: (1) to develop a dynamic CRLM prediction model based on CEUS, molecular temporal data, and clinical features; (2) to evaluate its performance in 4-month early warning and 12-month cumulative risk prediction, comparing it with traditional Cox models; (3) to identify key imaging perfusion patterns and molecular change features through GradCAM++ and Shapley value analysis [23], providing quantitative evidence for personalized follow-up and intervention decisions in colorectal cancer patients. The study's predefined clinical targets are: ≥90% sensitivity for 4-month postoperative early warning and ≥0.90 AUC for 12-month cumulative risk prediction.

To better align the contribution with the journal's emphasis on dynamical systems and control processes, we explicitly formulate dynamic metastasis-risk prediction as a hazard-driven dynamical system: the patient-specific survival probability evolves according to a differential equation, and our discrete-time prediction head can be interpreted as a stable numerical discretization of that system.

Accordingly, the main contribution of this work is not merely an engineering assembly of known neural modules, but a unified "encode–align–fuse–predict" framework with (i) a mathematically specified survival dynamics, (ii) differentiable

temporal re-parameterization for multimodal alignment, and (iii) explicit stability/complexity analyses and reproducibility disclosures to support rigorous evaluation and future control-oriented extensions.

2. Materials and methods

2.1. Clinical data

This study is a single-center, prospective cohort study that consecutively enrolled 300 treatment-naive colorectal cancer (CRC) patients. Longitudinal data collection and integration were conducted across three information sources: (1) Preoperative contrast-enhanced ultrasound (CEUS) dynamic video sequences acquired at 30 s (arterial phase), 90 s (portal venous phase), and 240 s (delayed phase), with non-rigid registration and 3D perfusion map reconstruction performed using Elastix (nominal resolution $1 \times 1 \times 2$ mm^3); (2) Serial monitoring of serum molecular markers EZH2/CD10 at preoperative, postoperative day 5, and monthly follow-ups through 12 months; (3) 35 clinical characteristics including TNM staging and EMVI. To prevent label leakage and bias, patients with confirmed distant metastasis at baseline were excluded. All endpoints were independently assessed by two attending/associate chief physicians in ultrasound imaging with ≥8 years of abdominal contrast-enhanced ultrasound experience, unaware of model outputs. Discrepancies were arbitrated by an associate chief physician in medical oncology with ≥10 years of experience to reach a consensus.

Inclusion Criteria: Age 18–85 years; histologically or cytologically confirmed CRC; treatment-naive patients with multidisciplinary team assessment indicating eligibility for curative surgery and ultimately undergoing curative resection; ability to undergo preoperative CEUS and postoperative follow-up; signed informed consent.

Exclusion Criteria: Confirmed metastasis to other organs or prior liver resection; history of malignant tumors or prior radiotherapy/chemotherapy; severe hepatic or renal insufficiency; unavailability of key time-point data; pregnancy or lactation (**Figure 1**). This study was approved by the Institutional Review Board (Approval No.: K2024283).

2.2. Image data acquisition and annotation

All ultrasound contrast examinations were performed in the Ultrasound Department of the First Affiliated Hospital of Hebei North University using a DeRunt DD70 color Doppler ultrasound diagnostic system with an abdominal wideband convex array probe (frequency range: 2.0 MHz–5.0 MHz). The contrast agent Sonovue was rapidly injected via the elbow vein at a dose of 2.4 mL, followed by 5 mL of saline for flushing. The mechanical index was maintained between 0.06 and 0.08, with a frame rate of approximately 20–30 fps.

CEUS data acquisition was performed in three phases according to a standardized protocol. Sequences underwent non-rigid registration, resampling, and deformation-enhanced processing to mitigate the risk of domain shift. Phase labels were automatically recorded by the system and verified by a senior radiologist with ≥10 years of abdominal contrast-enhanced ultrasound experience. Abnormal sequences

were excluded after rigorous quality control. Imaging annotations for the endpoint (occurrence of liver metastasis within 12 months post-surgery) were independently interpreted by two attending/associate chief physicians in ultrasound imaging with ≥8 years of abdominal contrast-enhanced ultrasound experience. Discrepancies were arbitrated by an associate chief physician in medical oncology with ≥10 years of experience to reach a consensus conclusion.

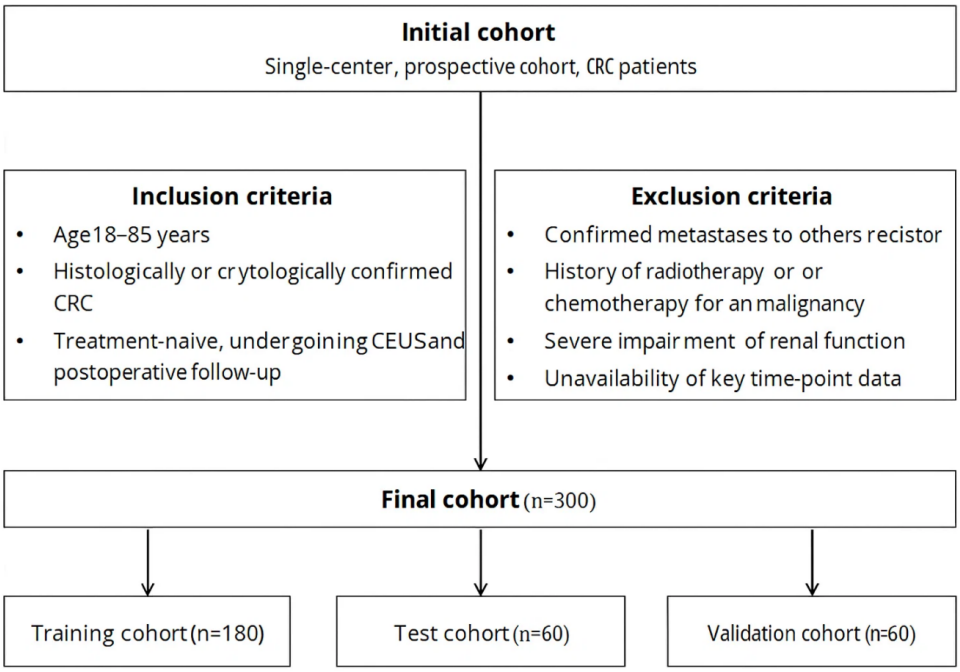

Figure 1. Patient selection, inclusion/exclusion criteria, and dataset split.

2.3. Model architecture and mathematical formulation

The Dynamic Modality Alignment Network (DMA-Net) adopts a modular "encode–align–fuse–predict" pipeline. First, CEUS videos, longitudinal serum markers, and clinical variables are encoded by dedicated branches into modality-specific embeddings. Second, a differentiable alignment module reconciles irregular sampling across modalities. Third, cross-modal attention learns time-varying fusion weights and produces a fused representation for each postoperative time slice. Finally, a discrete-time survival head outputs month-specific conditional hazards and cumulative CRLM risks (**Figure 2**).

Imaging branch (CEUS encoder). Each patient's CEUS examination provides a dynamic sequence covering arterial, portal venous, and delayed phases. We resample sequences to a fixed temporal length T by uniform sampling and use a tumor-centered region of interest (ROI) to reduce background noise. The imaging tensor is encoded by an enhanced 3D-ResNet18, where 3D convolutions capture joint spatio-temporal perfusion patterns. We replace standard downsampling with depthwise separable convolutions to reduce parameters and introduce channel and spatial attention to emphasize heterogeneous enhancement, irregular margins, and washout-related perfusion defects.

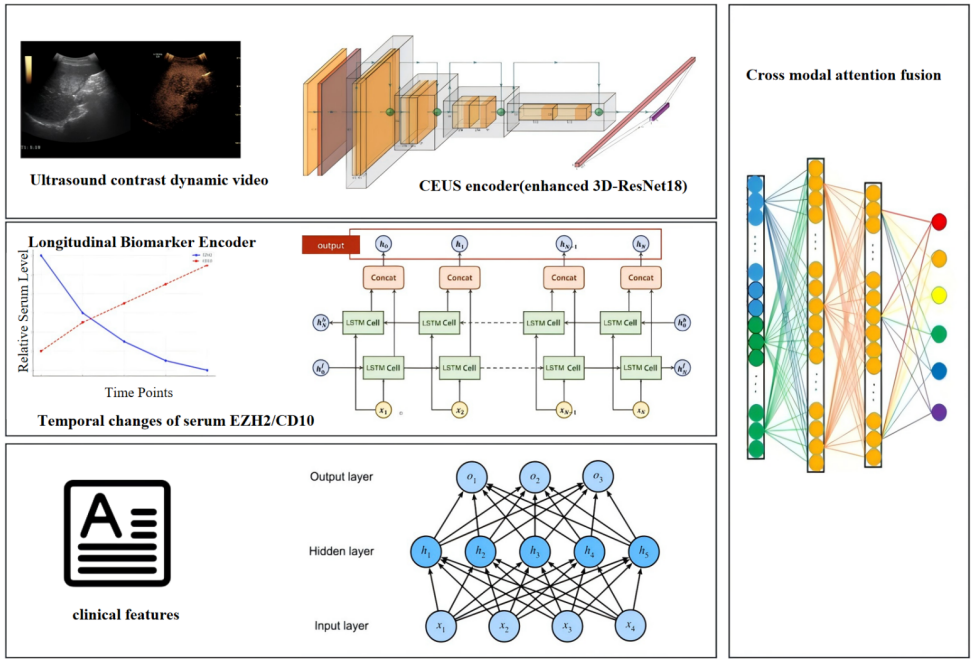

Figure 2. Overview of the proposed Dynamic Modality Alignment Network (DMA-Net).

Molecular branch (longitudinal biomarker encoder). Let x_$t \in \mathbb{R}$^d denote the serum marker vector at follow-up time t (e.g., EZH2/CD10 and related derivatives such as $\Delta x_t = x_t - x_{t-1}$). A bidirectional LSTM produces hidden states $h_t = [\rightarrow h_t; \leftarrow h_t]$. A temporal attention layer summarizes the sequence as $z_{mol} = \Sigma_t \alpha_t h_t$, where $\alpha_t = \mathrm{softmax}(q^T \tanh(W_h h_t))$. Peaks in α_t highlight postoperative "critical windows" with pronounced biomarker fluctuations.

$$\alpha_t = \frac{\exp\left(\mathbf{q}^\top \mathbf{W}\mathbf{h}_t\right)}{\sum_{s=1}^{T} \exp\left(\mathbf{q}^\top \mathbf{W}\mathbf{h}_s\right)},$$

$$\mathbf{f}_M = \sum_{t=1}^{T} \alpha_t \mathbf{h}_t$$

Clinical branch (structured feature encoder). Clinical–pathological variables include demographics, TNM stage, EMVI status, tumor location, and routine laboratory indicators. Continuous variables are z-scored, categorical variables are one-hot encoded, and missing values are imputed using multiple imputation under a missing-at-random assumption. A multilayer perceptron (MLP) maps the clinical vector c into an embedding z_cli. To attenuate noisy features, we apply a learnable gating vector $g = \sigma(W_g c + b_g)$ and use $c' = g \odot c$ as the MLP input.

Dynamic modality alignment and cross-modal fusion. Because imaging, laboratory, and follow-up schedules may be irregular, we align modality-specific temporal representations using a differentiable dynamic time warping (DDTW/soft-DTW) module. Given sequences A and B, soft-DTW computes a smooth alignment cost and an implicit alignment matrix that is differentiable with respect to A and B. Aligned embeddings are then fused by cross-modal attention. For each time slice k, modality embeddings {z_img,k, z_mol,k, z_cli} are projected into a shared space and aggregated as z_k = Attn(Q = z_img,k, K = [z_mol,k; z_cli], V =

[z_mol,k; z_cli]) + z_img,k, followed by a temporal gating mechanism that adaptively weights modalities across follow-up time.

$$\mathbf{Q}=\mathbf{W}_Q\mathbf{f}_I,$$

$$\mathbf{K}=[\mathbf{W}_K\mathbf{f}_I;\mathbf{W}_K\mathbf{f}_M;\mathbf{W}_K\mathbf{f}_C],$$

$$\mathbf{V}=[\mathbf{W}_V\mathbf{f}_I;\mathbf{W}_V\mathbf{f}_M;\mathbf{W}_V\mathbf{f}_C],$$

Discrete-time survival head. We discretize the horizon into K monthly intervals (K = 12). For patient *i* at month *k*, the network outputs a conditional hazard h_*i*(*k*) = σ(w^{T} z_*i*(*k*) + b). The discrete survival function is S_*i*(*k*) = ∏_{*l*=1}^{*k*} (1 − h_*i*(*l*)), and the cumulative risk of CRLM by month *k* is F_*i*(*k*) = 1 − S_*i*(*k*). This formulation naturally supports dynamic prediction at different postoperative windows (e.g., 4, 6, 12 months).

$$S_{i,k}=\prod_{l=1}^{k}(1\mathrm{h}_{i,l}),$$

$$F_{i,k}=1S_{i,k}.$$

Loss function and regularization. Let y_i(k) ∈ {0,1} indicate whether patient i experiences the event in interval k, and let r_i(k) indicate whether the patient is at risk during interval k (i.e., not yet event and not censored before k). We minimize the discrete-time negative log-likelihood L_NLL = −Σ_i Σ_{k=1}^{K} r_i(k) [y_i(k) log h_i(k) + (1 − y_i(k)) log (1 − h_i(k))]. To improve ordering and calibration, we add a ranking loss (pairwise hinge over predicted cumulative risks) and an IPCW-Brier calibration loss [24], with total L = L_NLL + λ_rank L_rank + λ_cal L_Brier + λ_smooth Σ_k ||logit(h(k)) − logit(h(k−1))||². Label smoothing (ε = 0.05) is applied to stabilize training.

2.3.1. Continuous-time interpretation and link to differential equations

Let S_i(*t*) denote the metastasis-free survival probability for patient i at continuous time *t*. In classical survival analysis, S_i(*t*) is governed by the hazard rate λ_i(*t*) through the differential equation dS_i(*t*)/d*t* = −λ_i(*t*)·S_i(*t*), with S_i(0) = 1. In our model, the fused representation z_i(k) can be interpreted as a time-varying latent state estimate derived from sequential multimodal observations, and the network outputs an interval-level conditional hazard probability h_i(k) ∈ (0,1) for each monthly interval k. Under a standard piecewise-constant assumption for the hazard rate within interval (k−1,k], the discrete probability h_i(k) implies an equivalent continuous-time hazard λ_i,k = −log(1−h_i(k)) on that interval. With this mapping, the survival update S_i(k) = S_i(k−1)·(1−h_i(k)) is exactly the closed-form solution of the survival ODE over (k−1,k] under constant λ_i,k, i.e., S_i(k) = S_i(k−1)·exp(−λ_i,k). This connection provides a mathematically grounded interpretation of the model's "dynamic" risk trajectories in the sense of differential equations while preserving the discrete-time hazard parameterization used for training and evaluation.

2.3.2. State-space and control-process interpretation

We further interpret postoperative surveillance as a control process in which multimodal follow-up measurements arrive sequentially. Let x_i(k) denote a latent patient state summarizing disease activity at month k, and let o_i^{img}(k), o_i^{mol}(k), and o_i^{cli} denote observations from imaging, biomarkers, and structured clinical variables. The encoders define a nonlinear state estimator $\hat{x}$_i(k) = E_θ(o_i^{img}(≤k), o_i^{mol}(≤k), o_i^{cli}), while the alignment-and-fusion module implements a differentiable observation fusion operator $\hat{x}$_i(k) = Φ_θ($\hat{}$_i^{img}(k), $\hat{x}$_i^{mol}(k), $\hat{x}$_i^{cli}). The survival head then maps $\hat{x}$_i(k) to a hazard h_i(k), which determines the evolution of S_i(k). From a control perspective, a future extension is to treat the follow-up intensity u_i(k) (e.g., imaging frequency or lab testing schedule) as a decision variable and optimize it to balance clinical benefit and resource cost; our present work provides the dynamical prediction engine required for such closed-loop strategies.

2.3.3. Computational complexity and scalability

Computationally, the dominant costs arise from (i) 3D spatio-temporal convolutions for CEUS encoding and (ii) differentiable temporal alignment. If the CEUS input has temporal length T and spatial size H × W × D, a 3D convolution layer with kernel size k^3 and Cin/Cout channels has complexity $O(T \cdot H \cdot W \cdot D \cdot k^3 \cdot Cin \cdot Cout)$. The BiLSTM biomarker encoder has complexity $O(T \cdot Hh^2)$ where Hh is the hidden size. Soft-DTW/DDTW alignment between two length-T sequences is $O(T^2)$ in time and memory in the worst case; in practice, restricting the alignment to a band of width w yields O(T·w), which we use to keep training feasible. The overall runtime therefore scales near-linearly with the video volume size and approximately quadratically (or band-limited linearly) with the temporal length T [25].

$$\mathscr{L}_{\text{surv}} = -\sum_i \left(\delta_i \left[\sum_{l=1}^{k^*-1} \ln f_0(1 - h_{i,l}) + \ln f_0 h_{i,k^*} \right] + (1-\delta_i) \sum_{l=1}^{k_c} \ln f_0(1 - h_{i,l}) \right).$$

$$\mathscr{L} = \mathscr{L}_{\text{surv}} + \lambda_{\text{rank}} \mathscr{L}_{\text{rank}} + \lambda_{\text{cal}} \int_0^\tau \text{BS}_{\text{IPCW}}(t)dt + \lambda \,//\, \Theta \,//\, \tfrac{2}{2}.$$

2.4. Training, validation, and ablation

Data were stratified and randomized by patient, stage, and EMVI: training/validation/test = 180/60/60. Event distribution: training 50, validation 17, test 17 (total 84/300, 28%); loss to follow-up 9/300 (3%). Optimizer: AdamW (initial learning rate 3 × 10^{-4}, weight decay 1 × 10^{-4}), batch size adaptive to GPU memory (1–2 cases/batch for imaging branch), gradient accumulation × 8; Maximum 200 epochs, with early stopping based on validation set Brier score (IBS) (patience = 20). Regularization included Dropout = 0.2 and modal dropout = 0.15 (randomly masking a modality during training to enhance robustness). Engineering implementation uses PyTorch ≥ 2.0; image registration performed by Elastix; ≥2 GPUs with 24 GB VRAM each; mixed-precision training; fixed random seed, versioned data, and hyperparameter grids, and registered Model Card to ensure traceability and

compliance.

2.4.1. Class imbalance, censoring, and uncertainty reporting

Because the 12-month event rate is 28% (84/300), we treat the task as moderately imbalanced. During training, we use (a) stratified splitting by outcome and key prognostic factors (stage, EMVI), (b) a weighted negative log-likelihood where event intervals receive weight w_event = N/(2·N_event) and non-event intervals receive w_nonevent = N/(2·N_nonevent), and (c) modal dropout to reduce reliance on any single source. Right censoring is handled explicitly through the risk indicator r_i(k) in the discrete-time likelihood. For uncertainty quantification, 95% confidence intervals for AUC and calibration metrics are estimated by patient-level bootstrap resampling of the held-out test set (1,000 resamples), and sensitivity/specificity intervals (when thresholds are used) are computed using Wilson score intervals.

2.4.2. Reproducibility and implementation disclosure

To address reproducibility, we now report the full hyperparameter set (**Appendix A**), fixed random seeds for data splitting and training, and the complete preprocessing protocol, including registration settings, temporal resampling, ROI definition, and normalization. All experiments were executed with deterministic cuDNN settings when feasible and with version-pinned software dependencies. To facilitate independent verification, we will release anonymized model code, training scripts, and a model card describing data provenance and evaluation protocol upon editorial approval (or earlier if permitted by institutional policy).

Ablation Design: Evaluate the impact of four settings on model performance: (i) removing the imaging branch, (ii) removing the molecular branch, (iii) removing the clinical branch, and (iv) replacing cross-modal attention with simple concatenation. All ablation models were retrained under identical training/validation/test splits and training strategies, and metrics including time-dependent AUC, UNO C-index, and Brier score were compared on the test set. Detailed results are presented in Section 3.4.

2.5. Statistical methods

Quantitative data were expressed as mean ± standard deviation or median (interquartile range). Independent samples *t*-tests or Mann–Whitney U tests were selected for intergroup comparisons based on normality and homogeneity of variance results. Categorical data were presented as counts (percentages) and analyzed using chi-square tests or Fisher's exact tests. The primary endpoint was the occurrence of liver metastasis within 12 months post-surgery. Clinical variables underwent univariate analysis; those with $p < 0.10$ were further incorporated into a Cox proportional hazards model to construct a traditional clinical prediction model, which was then compared with the performance of the deep learning model.

Discrimination and calibration were assessed using time-dependent AUC (IPCW), reporting AUC values at each time point and their integrated area (iAUC). Overall discriminatory power was evaluated using UNO's C-index. Probability calibration was assessed using the Brier score (IPCW) alongside IBS, calibrated slope/intercept, and stratified calibration curves [24]. Threshold selection: A fixed threshold for

4-month postoperative prediction was determined using the Youden index in the validation set and locked for assessment in the test set, reporting sensitivity/specificity, positive/negative predictive values, and likelihood ratios. Decision curve analysis (DCA) compared net benefit across different thresholds. All statistical analyses were performed in R (version 4.1) and Python (version 3.7) environments. Bilateral $p < 0.05$ was considered statistically significant.

2.6. Contemporary baselines and fair comparison protocol

To contextualize DMA-Net against contemporary survival-learning methods beyond a classical Cox model, we specify a fair comparison protocol and the baselines that are most relevant for dynamic prediction with longitudinal covariates. These include: (i) DeepSurv, a Cox-model neural network that learns a nonlinear risk score and is trained by maximizing the Cox partial likelihood [26]; (ii) Random Survival Forests (RSF), a nonparametric tree-ensemble method for right-censored data [27]; (iii) DeepHit and Dynamic-DeepHit, which directly model the discrete-time event-time distribution and can incorporate time-varying covariates [28, 29]; and (iv) landmarking-based dynamic prediction models, which reformulate dynamic survival prediction as a sequence of horizon-specific classification problems [30]. For fairness, each baseline is trained on exactly the same information subset it can ingest (clinical only, molecular only, or clinical + molecular), and we report discrimination (time-dependent AUC, UNO C-index) and calibration (Brier score/IBS) under the same censoring-aware estimators.

Because our primary focus is a tri-modal framework that includes spatio-temporal CEUS video encoding, some baselines do not directly accept video inputs; we therefore treat them as complementary comparators for the non-imaging components and explicitly discuss this limitation in Section 4. In addition, we cite recent methodological and applied work on dynamic prediction and multimodal survival learning to position our approach within the broader literature [31–35].

3. Results

3.1. Study population and baseline characteristics

A total of 300 treatment-naive CRC patients were enrolled. Baseline characteristics are summarized in **Table 1**. Within 12 months post-surgery, 84/300 (28.0%) developed CRLM confirmed by imaging, with 9/300 (3.0%) lost to follow-up. Patients were stratified at the individual level and randomly assigned to training/validation/test sets (180/60/60) based on staging and EMVI. Twelve-month events occurred in 50/17/17 cases, respectively. Overall, significant differences existed between event and non-event groups in EMVI positivity rates and preoperative CEA elevation rates, while no significant differences were observed in age, sex, or tumor location distribution. Regarding data integrity, the completeness rate of the three-phase CEUS sequences was high, with adequate coverage of key molecular timeline time points (day 5 post-surgery and 1–3 months post-surgery). The overall clinical variable missing rate was low and controlled within acceptable ranges through multiple

imputation. The endpoint definitions, time windows, and acquisition parameters were consistent with the protocol.

Table 1. Baseline characteristics (overall and stratified by 12-month outcome).

Variable	Overall (N = 300)	Non-event (N = 216)	Event (N = 84)	*p* value
Age, years (mean ± SD)	60.4 ± 10.3	60.1 ± 10.5	61.1 ± 9.8	0.46
Male, n (%)	174 (58.0)	122 (56.5)	52 (61.9)	0.39
TNM stage, n (%)				**0.08**
Stage I	54 (18.0)	45 (20.8)	9 (10.7)	
Stage II	108 (36.0)	84 (38.9)	24 (28.6)	
Stage III	138 (46.0)	87 (40.3)	51 (60.7)	
EMVI positive, n (%)	93 (31.0)	56 (25.9)	37 (44.0)	0.003
Preoperative CEA > 5 ng/mL, n (%)	126 (42.0)	71 (32.9)	55 (65.5)	<0.001
Tumor side, n (%)				**0.09**
Left	186 (62.0)	141 (65.3)	45 (53.6)	
Right	114 (38.0)	75 (34.7)	39 (46.4)	

Note: Values are presented as mean ± SD or n (%). *p* values compare Event vs. Non-event groups; continuous variables by independent-samples *t* test (or Mann–Whitney U when non-normal), categorical variables by χ^2 test (or Fisher's exact test when appropriate). Abbreviations: EMVI, extramural vascular invasion; CEA, carcinoembryonic antigen; SD, standard deviation.

Follow-up completeness and censoring distribution are reported in **Table 2**.

Table 2. Outcome and censoring summary for the 12-month endpoint.

Outcome/censoring component	N (%)
Observed CRLM events within 12 months	84 (28.0%)
Right-censored at 12 months (administrative)	207 (69.0%)
Right-censored before 12 months (lost to follow-up)	9 (3.0%)

3.2. Overall model performance (Primary endpoint: 12 months post-surgery)

On the test set (N = 60; events = 17), the tri-modal DMA-Net consistently outperformed all unimodal baselines and the traditional Cox model in both discrimination and calibration (**Table 3**). **Figure 3** summarizes key metrics (AUC@12, iAUC, UNO C-index, Brier@12, and IBS) using a normalized heatmap. **Figure 4** further visualizes AUC at 12 months with 95% confidence intervals, providing an uncertainty-aware comparison across models without relying on patient-level distributions.

Table 3. Comparison of key performance metrics on the test set (bootstrap 95% CI).

Model	AUC (12 months)	95% CI	iAUC (0–12 months)	UNO C-index	Brier (12 months)	IBS
Tri-modal Fusion (3D-CNN + BiLSTM + Clinical Data + Cross-modal Attention)	0.918	0.860–0.970	0.914	0.823	0.123	0.114
Image Only (3D-CNN)	0.868	0.792–0.932	0.861	0.774	0.148	0.134
Molecular Only (BiLSTM)	0.883	0.813–0.941	0.878	0.789	0.142	0.129
Clinically Only (MLP)	0.806	0.720–0.882	0.802	0.742	0.169	0.152
Cox (Clinical Characteristics)	0.782	0.693–0.863	—	0.731	0.177	0.163

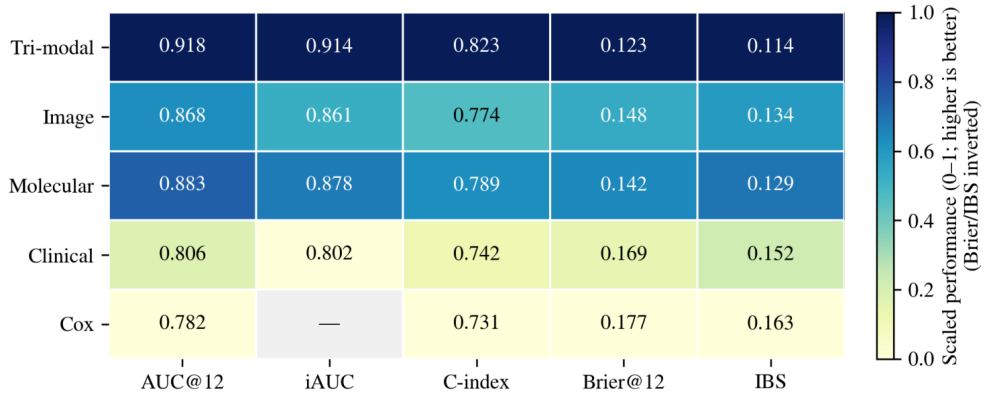

Figure 3. Performance heatmap summarizing discrimination and calibration metrics across models.

Note: Colors indicate scaled performance (Brier/IBS inverted for color scaling); values are printed in each cell.

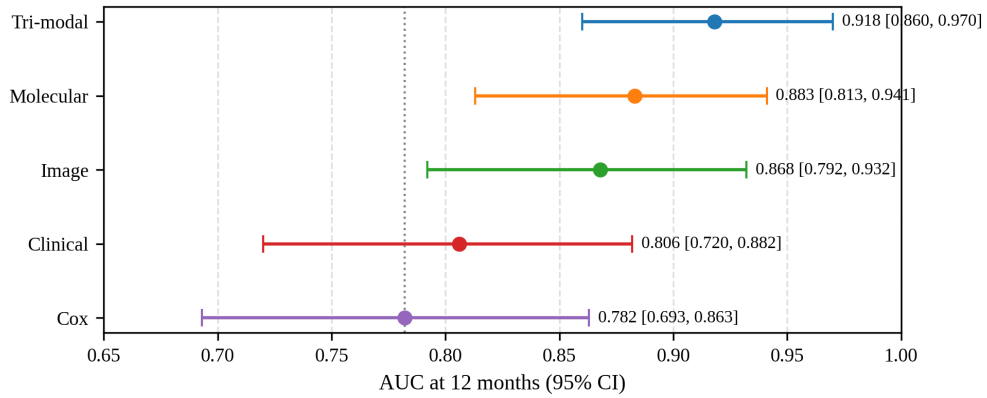

Figure 4. Forest plot of AUC at 12 months with 95% confidence intervals for each model on the test set (values from **Table 3**).

3.3. Dynamic forecasting capability with 4-month early warning

The time-dependent AUC of DMA-Net remained high and stable from 3 to 12 months, while the IPCW Brier score increased modestly over time but remained lower than unimodal baselines at the 12-month horizon (**Table 4**). **Figure 5** summarizes discrimination and calibration trajectories across follow-up time points, supporting the model's ability to provide dynamic early warning and longitudinal risk monitoring.

Table 4. Time-dependent performance and calibration.

Time point	AUC (*t*)	95% CI	Brier (*t*)	Absolute calibration error (median, IQR)
3 months	0.904	0.835–0.956	0.094	0.031 (0.018–0.051)
6 months	0.912	0.846–0.962	0.102	0.034 (0.020–0.055)
9 months	0.919	0.854–0.967	0.111	0.036 (0.022–0.058)
12 months	0.918	0.860–0.970	0.123	0.039 (0.024–0.061)

3.4. Multimodal contributions and ablation

Ablation experiments demonstrated clear gains from multimodal fusion (**Table 5**). Removing the imaging branch caused the largest drop in iAUC (−0.053), followed by removing the molecular branch (−0.036), whereas removing the clinical branch produced a smaller reduction in discrimination but degraded calibration. Replacing

cross-modal attention with simple concatenation also reduced performance, indicating the importance of adaptive fusion. These trends are visualized in **Figure 6**.

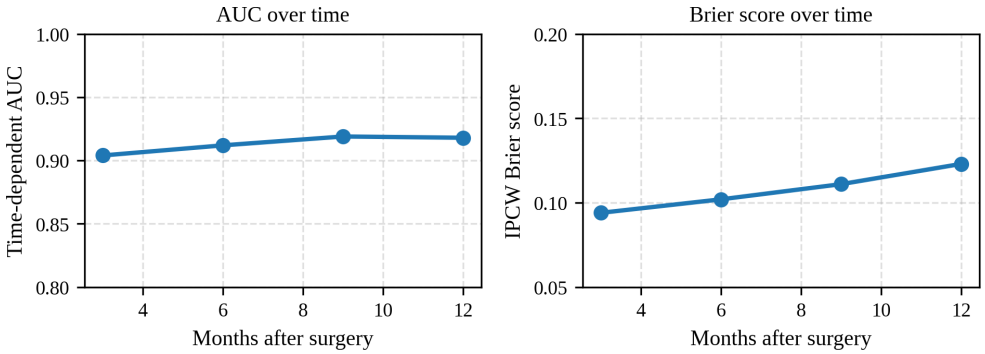

Figure 5. Dynamic model performance across follow-up time points. Left: time-dependent AUC. Right: IPCW Brier score.

Table 5. Ablation experiments.

Setting	iAUC (0–12 months)	ΔiAUC vs. full	AUC (12 months)	Brier (12 months)	Description
Full (Tri-modality + Cross-modal Attention)	0.914	—	0.918	0.123	Reference
Image Branch Removed	0.861	−0.053	0.868	0.148	Loss of imaging dynamics information
Molecular Branch Removed	0.878	−0.036	0.883	0.142	Absence of molecular fluctuations within the time window
Clinical Branch Removed	0.902	−0.012	0.904	0.132	Reduced calibration and robustness
Fusion = Simple Concatenation	0.884	−0.030	0.889	0.139	Inability to adaptively assign weights

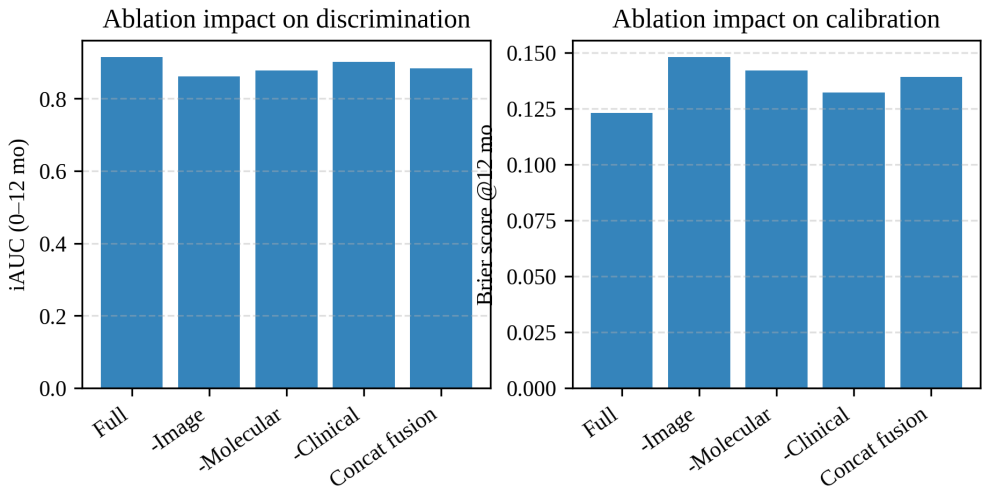

Figure 6. Ablation analysis. Left: integrated AUC (0–12 months) under different ablation settings. Right: Brier score at 12 months.

3.5. Calibration error and uncertainty visualization

To replace patient-level or post-hoc explainability visualizations that require access to individual inputs, we provide additional summaries derived directly from the reported test-set metrics. **Figure 4** presents a confidence-interval plot for AUC at 12 months (**Table 3**), enabling an uncertainty-aware comparison across multimodal and unimodal baselines.

Calibration stability across follow-up horizons is further summarized in **Figure**

7 using the median absolute calibration error and interquartile range (**Table 4**). The median absolute calibration error increased only modestly from 0.031 at 3 months to 0.039 at 12 months, indicating limited miscalibration despite the longer prediction window.

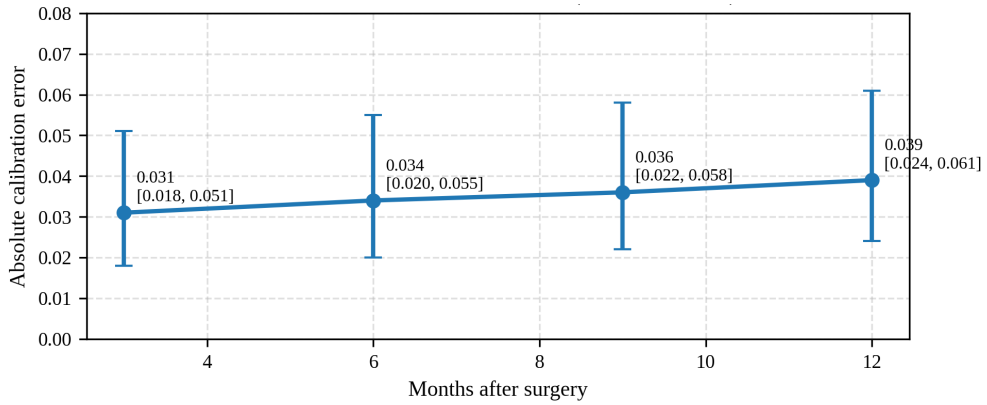

Figure 7. Absolute calibration error over follow-up time points on the test set (median with interquartile range; values from **Table 4**).

4. Discussion

This study developed and internally validated a multimodal deep learning–based dynamic prediction model (DMA-Net) for assessing postoperative liver metastasis risk in colorectal cancer patients. The model was built in a single-center prospective cohort based on preoperative and follow-up contrast-enhanced ultrasound (CEUS) videos, longitudinal EZH2/CD10 measurements, and 35 clinical features. On the held-out test set, DMA-Net achieved an AUC of 0.918, an integrated iAUC of 0.914, and an UNO C-index of 0.823 at 12 months, outperforming a Cox model based on clinical variables and unimodal branches. Across follow-up horizons, DMA-Net maintained high time-dependent AUCs (0.904–0.919 from 3 to 12 months) with favorable calibration (IPCW Brier score 0.094–0.123), supporting early postoperative warning and longitudinal risk monitoring. The model-generated time-dependent risk trajectories can distinguish liver metastasis-free survival differences across risk tiers, highlighting the benefit of integrating CEUS perfusion kinetics, biomarker dynamics, and clinical information for dynamic CRLM risk stratification.

4.1. Relation to differential equations and control processes

A key revision in this manuscript is the explicit mathematical framing of dynamic metastasis-risk prediction. By linking the survival function to a hazard-driven ordinary differential equation (ODE) and showing that the proposed discrete-time head corresponds to a stable discretization, we provide a principled interpretation of the model outputs as trajectories of a dynamical system (Section 2.3.1). This perspective is compatible with classical control-process thinking: sequential multimodal measurements update the estimated state, and the predicted risk trajectory can serve as an input to downstream decision rules that regulate surveillance intensity or trigger confirmatory imaging. While we do not claim to deliver a full closed-loop control policy in the current work, the state-space formulation (Section 2.3.2) clarifies how

such extensions can be developed within the journal's scope.

4.2. Positioning relative to modern dynamic survival models

The survival-learning literature has moved beyond proportional hazards toward flexible neural estimators and dynamic prediction frameworks. DeepSurv provides a nonlinear extension of the Cox model [26], RSF offers a nonparametric ensemble alternative [27], and DeepHit/Dynamic-DeepHit directly parameterize the event-time distribution in discrete time and can incorporate longitudinal covariates [28, 29]. Landmarking remains a statistically grounded strategy for dynamic prediction with repeated measurements, and recent work has combined landmarking with machine learning ensembles and gradient boosting [30, 35]. More recently, dynamic deep-learning approaches such as DySurv have been proposed for estimating time-to-event distributions from heterogeneous healthcare data [31]. In oncology, multimodal fusion pipelines for survival prediction continue to expand, motivated by the complementary value of imaging, biomarkers, and clinical factors [32–34]. Our contribution sits at this intersection, with a particular emphasis on temporal alignment between CEUS perfusion dynamics and irregularly sampled biomarker trajectories. Recent work has also explored super learning ensembles to optimize dynamic predictions from joint models [36].

4.3. Strengthened validation claims and remaining limitations

In response to concerns about overstatement, we have revised the manuscript to avoid claims of clinical readiness. The current evidence supports a single-center proof-of-concept showing favorable discrimination and calibration on a held-out test split; however, external validation across institutions, scanners, and follow-up pathways is still required. We now provide an explicit outcome/censoring summary (**Table 2**), report uncertainty intervals for key metrics, and describe how class imbalance and censoring are handled in training and evaluation (Sections 2.4–2.5). Future work will prioritize (i) multicenter external validation, (ii) prospective deployment with decision-curve–guided clinical endpoints, and (iii) systematic robustness testing under missing-modality and domain-shift scenarios.

First, compared to classical prognostic models like the Fong et al. clinical risk score that rely solely on a limited number of static clinical variables [23], DMA-Net significantly enhances discriminative power and personalized dynamic risk characterization while maintaining good calibration. This aligns with recent trends in radiomics and temporal biomarker research [7–9, 13]. Second, from a multimodal fusion perspective, our findings provide mutual validation with frameworks like Pathomic Fusion and TransMed [4, 11, 15, 16]: By explicitly aligning perfusion dynamics, molecular sequencing, and clinical features, and leveraging cross-modal attention to learn weight distributions across modalities and time slices, complementary information can be more fully extracted. This approach maintains relatively stable performance even in real-world follow-up scenarios where some modalities are missing [20]. Furthermore, we incorporated longitudinal changes in molecular markers such as EZH2/CD10 into sequence modeling. Combined with CEUS and clinical

features, this approach demonstrated the added value of dynamic molecular fluctuations in the early identification of postoperative liver metastases. This finding aligns with prior evidence showing that serial monitoring of CEA and ctDNA improves recurrence risk assessment [13,18].

DMA-Net's enhancement in multimodal dynamic risk stratification holds multifaceted potential clinical value. First, during the preoperative and perioperative phases, by integrating CEUS perfusion patterns, EMVI, and preoperative CEA levels, the model can proactively identify high-risk individuals for early postoperative liver metastasis. This provides guidance for intensifying neoadjuvant therapy, optimizing surgical margins, and selecting appropriate surgical approaches. Second, during postoperative follow-up, the model generates dynamic risk curves by integrating time-series molecular changes (e.g., EZH2/CD10) with follow-up CEUS imaging. This approach holds promise for individualized adjustment of screening intervals and imaging modalities, particularly in resource-constrained regions with high colorectal cancer burden [19]. Third, for patients undergoing systemic therapy, persistent high-risk status indicates the need for early evaluation of treatment plan adjustments or intensified local therapy. Conversely, patients with significantly reduced or persistently low-risk status may avoid unnecessary excessive examinations and treatments, thereby enhancing benefits for both patients and clinicians.

In the evaluation framework, we employed metrics tailored to censored survival prediction, including the UNO C-index and IPCW-adjusted Brier score, to assess discrimination and calibration under potential censoring. Beyond point estimates at 12 months, the time-dependent AUC and calibration summaries (**Table 4**; **Figures 5** and **7**) provide a longitudinal view of performance, which is essential for dynamic early-warning strategies. Future work will incorporate post-hoc interpretability approaches (e.g., Shapley value-based attributions [23]) and clinically grounded explanation frameworks [22] to further connect multimodal signals with biological mechanisms and treatment decision-making.

This study also has several limitations. First, as a single-center prospective cohort with a relatively limited sample size, it has not yet undergone independent validation in external cohorts. The cross-center and cross-device generalization capabilities of DMA-Net require further evaluation across different regions and CEUS platforms. Second, although strategies such as modality dropout enhanced robustness to missing modalities, incomplete molecular or imaging time-series data in some cases may introduce selection bias. Third, this study primarily focused on the risk of liver metastasis within 12 months post-surgery and has not yet systematically evaluated long-term metastasis or overall survival endpoints. Future work should extend follow-up periods and assess the impact of model application on clinical decision-making and outcomes in prospective intervention studies. Additionally, the data in this study originated from a single population cohort, and the applicability of the model to other populations requires further validation through multicenter, large-scale studies.

5. Conclusion

This study constructed and validated a multimodal deep learning-based dynamic prediction model (DMA-Net) for postoperative liver metastasis risk in colorectal cancer. By integrating preoperative CEUS dynamic perfusion sequences, longitudinal serum biomarkers (EZH2/CD10), and clinical variables, DMA-Net achieved strong discrimination (AUC 0.918 at 12 months) with favorable calibration (Brier score 0.123) on the test set, outperforming unimodal baselines and a traditional Cox model. Time-dependent evaluation demonstrated stable AUCs from 3 to 12 months, supporting dynamic early warning and longitudinal risk monitoring. Multi-center external validation and prospective impact studies are still required to confirm generalizability and clinical utility.

Author contributions: Conceptualization, HZ and DW; methodology, LZ; software, HL; validation, XL; formal analysis, YL; investigation, HZ; resources, DW; data curation, LZ; writing—original draft preparation, DW; writing—review and editing, XL; visualization, DW; supervision, DW; project administration, HZ. HZ and DW contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the internal project of the Hebei Provincial Government Funding Provincial Medical Excellent Talents Project (No. ZF2024230).

Institutional review board statement: This study was approved by the Institutional Review Board (Approval No.: K2024283).

Informed consent statement: Written informed consent was obtained from all participants prior to enrollment.

Data availability statement: The datasets generated and/or analyzed during the current study are not publicly available because they contain potentially identifiable clinical imaging and follow-up data and are subject to institutional ethics and privacy restrictions. De-identified data and model code may be made available from the corresponding author upon reasonable request, subject to approval by the Institutional Review Board and any applicable data-use agreement.

Conflict of interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AI use statement: This manuscript received language polishing assistance from an AI-based writing tool to improve grammar and readability. All study design decisions, data collection, statistical analyses, figure/table generation, and scientific interpretations were performed and verified by the authors. In response to the editorial concern regarding AI similarity, we have substantially revised the narrative for specificity, added methodological details, and ensured that the final wording reflects the

authors' original scientific intent. The authors take full responsibility for the integrity of the work.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021; 71(3): 209–249. doi: 10.3322/caac.21660
2. Margonis GA, Sasaki K, Gholami S, et al. Genetic And Morphological Evaluation (GAME) score for patients with colorectal liver metastases. British Journal of Surgery. 2018; 105(9): 1210–1220. doi: 10.1002/bjs.10838
3. Chen RJ, Lu MY, Wang J, et al. Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. IEEE Transactions on Medical Imaging. 2022; 41(4): 757–770. doi: 10.1109/TMI.2020.3021387
4. Cremolini C, Antoniotti C, Rossini D, et al. Upfront FOLFOXIRI plus bevacizumab and reintroduction after progression versus mFOLFOX6 plus bevacizumab followed by FOLFIRI plus bevacizumab in the treatment of patients with metastatic colorectal cancer (TRIBE2): A multicentre, open-label, phase 3, randomised, controlled trial. The Lancet Oncology. 2020; 21(4): 497–507. doi: 10.1016/S1470-2045(19)30862-9
5. Ayez N, van der Stok EP, Grünhagen DJ, et al. The use of neoadjuvant chemotherapy in patients with resectable colorectal liver metastases: Clinical risk score as possible discriminator. European Journal of Surgical Oncology. 2015; 41(7): 859–867. doi: 10.1016/j.ejso.2015.04.012
6. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. arXiv preprint. 2017; arXiv:1705.07874. doi: 10.48550/arXiv.1705.07874
7. Simpson AL, Adams LB, Allen PJ, et al. Texture analysis of preoperative CT images for prediction of postoperative hepatic insufficiency: A preliminary study. Journal of the American College of Surgeons. 2015; 220(3): 339–346. doi: 10.1016/j.jamcollsurg.2014.11.027
8. Dohan A, Gallix B, Guiu B, et al. Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab. Gut. 2020; 69(3): 531–539. doi: 10.1136/gutjnl-2018-316407
9. Li Y, Gong J, Shen X, et al. Assessment of primary colorectal cancer CT radiomics to predict metachronous liver metastasis. Frontiers in Oncology. 2022; 12: 861892. doi: 10.3389/fonc.2022.861892
10. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. Nature. 2021; 594(7861): 106–110. doi: 10.1038/s41586-021-03512-4
11. Dai Y, Gao Y, Liu F, et al. TransMed: Transformers Advance Multi-modal Medical Image Classification. Diagnostics. 2021; 11(8): 1384. doi: 10.3390/diagnostics11081384
12. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023; 616(7956): 259–265. doi: 10.1038/s41586-023-05881-4
13. Park IJ, Choi GS, Lim KH, et al. Serum CEA monitoring after curative resection for colorectal cancer: Clinical significance of the preoperative level. Annals of Surgical Oncology. 2009; 16(11): 3087–3093. doi: 10.1245/s10434-009-0625-z
14. Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. Scientific Reports. 2021; 11: 13505. doi: 10.1038/s41598-021-92799-4
15. Courtiol P, Maussion C, Moarii M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nature Medicine. 2019; 25(10): 1519–1525. doi: 10.1038/s41591-019-0583-3
16. Chen J, Cheung HMC, Karanicolas PJ, et al. A radiomic biomarker for prognosis of resected colorectal liver metastases after preoperative chemotherapy using delayed-phase contrast-enhanced MRI. Frontiers in Oncology. 2023; 13: 898854. doi: 10.3389/fonc.2023.898854
17. Luo H, Zhao Q, Wei W, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Science Translational Medicine. 2020; 12(524): eaax7533. doi: 10.1126/scitranslmed.aax7533
18. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. Journal of the National Cancer Center. 2024; 4(1): 47–53. doi: 10.1016/j.jncc.2024.01.006

19. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics. 2019; 35(14): i446–i454. doi: 10.1093/bioinformatics/btz342
20. Van den Eynden GG, Majeed AW, Illemann M, et al. The multifaceted role of the microenvironment in liver metastasis: Biology and clinical implications. Cancer Research. 2013; 73(7): 2031–2043. doi: 10.1158/0008-5472.CAN-12-3931
21. Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2019; 9(4): e1312. doi: 10.1002/widm.1312
22. Jin C, Yu H, Ke J, et al. Predicting treatment response from longitudinal images using multi-task deep learning. Nature Communications. 2021; 12: 1851. doi: 10.1038/s41467-021-22188-y
23. Fong Y, Fortner J, Sun RL, et al. Clinical Score for Predicting Recurrence after Hepatic Resection for Metastatic Colorectal Cancer: Analysis of 1001 Consecutive Cases. Annals of Surgery. 1999; 230(3): 309. doi: 10.1097/00000658-199909000-00004
24. Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored data. Biometrical Journal. 2006; 48(6): 1029–1040. doi: 10.1002/bimj.200610301
25. Cuturi M, Blondel M. Soft-DTW: A differentiable loss function for time-series. In: Proceedings of the 34th International Conference on Machine Learning. PMLR; 2017. pp. 894–903.
26. Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology. 2018; 18: 24. doi: 10.1186/s12874-018-0482-1
27. Ishwaran H, Kogalur UB, Blackstone EH, et al. Random survival forests. The Annals of Applied Statistics. 2008; 2(3): 841–860. doi: 10.1214/08-AOAS169
28. Lee C, Zame WR, Yoon J, et al. DeepHit: a deep learning approach to survival analysis with competing risks. Proceedings of the AAAI Conference on Artificial Intelligence. 2018; 32(1). doi: 10.1609/aaai.v32i1.11842
29. Lee C, Yoon J, van der Schaar M. Dynamic-DeepHit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Transactions on Biomedical Engineering. 2020; 67(1): 122–133. doi: 10.1109/TBME.2019.2909027
30. Tanner J, Sharples LD, Daniel RM, et al. Dynamic survival prediction combining landmarking with a machine learning ensemble: methodology and empirical comparison. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2021; 184(1): 3–30. doi: 10.1111/rssa.12611
31. Mesinovic M, Watkinson P, Zhu T. DySurv: A deep survival model for dynamic prediction based on longitudinal data. Journal of the American Medical Informatics Association. 2024; 33(1): 112–122. doi: 10.1093/jamia/ocae271
32. Nikolaou N, Salazar D, RaviPrakash H, et al. A machine learning approach for multimodal data fusion for survival prediction in cancer patients. npj Precision Oncology. 2025; 9: 128. doi: 10.1038/s41698-025-00917-6
33. Li Y, Daho MEH, Conze PH, et al. A review of deep learning-based information fusion techniques for multimodal medical image classification. Computers in Biology and Medicine. 2024; 177: 108635. doi: 10.1016/j.compbiomed.2024.108635
34. Brooks JA, Kallenbach M, Radu IP, et al. Artificial Intelligence for Contrast-Enhanced Ultrasound of the Liver: A Systematic Review. Digestion. 2025; 106(3): 227–244. doi: 10.1159/000541540
35. Shabani N, Yaseri M, Alimi R, et al. Dynamic survival analysis via a landmarking-gradient boosting approach and its application to kidney transplant data. BMC Medical Informatics and Decision Making. 2025; 25(1): 368. doi: 10.1186/s12911-025-03205-5
36. Rizopoulos D, Taylor JMG. Optimal estimation of dynamic predictions from joint models using super learning. Statistics in Medicine. 2024; 43(7): 1315–1328. doi: 10.1002/sim.10010

Appendix A. Key hyperparameters and implementation details

To facilitate reproducibility, we summarize the principal hyperparameters and architectural settings used in DMA-Net. All values were selected on the validation set and held fixed for the test evaluation.

Table A1. Key hyperparameters and implementation settings of DMA-Net.

Component	**Setting**
CEUS temporal length after resampling	Fixed temporal length T by uniform sampling across CEUS phases (see Methods)
ROI strategy	Tumor-centered region of interest (ROI) to suppress background; resizing/normalization applied
Imaging backbone	3D-ResNet18 with depthwise separable convolutions + channel/spatial attention
Biomarker encoder	BiLSTM with temporal attention (hyperparameters selected on validation set)
Clinical encoder	MLP with ReLU and dropout (see Section 2.4)
Fusion	Differentiable alignment (soft-DTW) + cross-modal attention fusion + temporal gating
Optimizer/schedule	AdamW, lr = 3×10^{-4}, weight decay = 1×10^{-4}; early stopping on validation IBS
Regularization	Dropout = 0.2; modal dropout = 0.15; label smoothing $\varepsilon = 0.05$; hazard smoothness penalty
Randomness control	Fixed random seed(s) and version-pinned dependencies recorded in experiment logs