Appendix 1

Treatment regimens

All patients received radiotherapy delivered exclusively via intensity-modulated radiotherapy (IMRT) techniques. For those undergoing hyperfractionated radiotherapy, the regimen consisted of 1.5 Gy per fraction administered BID, to a total dose of 45–60 Gy. For patients receiving conventional fractionation, the treatment delivered 1.8–2 Gy per fraction QD, to a total dose of 54–66 Gy. All patients were positioned under large-bore CT, and the CT positioning images were transferred to the Varian Eclipse planning system. Daily image-guided radiotherapy (IGRT) was implemented to ensure treatment accuracy. Each patient was positioned by a radiation oncologist and a radiation physicist. The chemotherapy regimen primarily included intravenous administration of etoposide and cisplatin (EP regimen) or etoposide and carboplatin (EC regimen). The EP regimen consisted of etoposide (100 mg/m^2, intravenous infusion, days 1–3) + cisplatin (75 mg/m^2, intravenous infusion, day 1), repeated every 3–4 weeks. The EC regimen includes etoposide (100 mg/m^2, intravenous infusion, days 1–3) + carboplatin (AUC =5–6, intravenous infusion, day 1), repeated every 3 weeks. The chemotherapy regimen is administered over 4–6 cycles. Concurrent or sequential thoracic radiotherapy is administered during chemotherapy.

Appendix 2

VISTA3D segmentation DL model

This paper employs VISTA3D—a unified segmentation foundation model for 3D medical imaging—based on the SegResNet backbone and sliding-window inference. By integrating 3D super-voxel distillation (SAM) knowledge and a four-stage training approach, it achieves automatic segmentation of 127 structural classes, 3D interactive refinement, and zero-shot segmentation. Its inference speed is 3 to 4 times faster than TotalSegmentator. It demonstrates outstanding performance in few-shot fine-tuning, providing a comprehensive workflow solution for clinical 3D image segmentation. The VISTA3D model was fine-tuned on data from Center A, while data from Center B continued to serve as external test sets for the object segmentation task. The preprocessing and training workflow was designed to enhance data validity and optimize model performance: First, key regions were preserved through foreground cropping. Image intensities were then normalized to the range [0, 1] from [–110, 190]. to the range [0, 1]. Simultaneously, multi-modal data augmentation techniques are introduced, including random cropping by label category, random scaling, Gaussian smoothing, intensity scaling and offset, and Gaussian noise addition, to enhance the model's generalization capability.

Model and training hyperparameters were tailored to the segmentation task requirements: The optimizer employed AdamW with weight decay, an initial learning rate of 5e–5, and a WarmupCosineSchedule learning rate scheduler. The loss function utilized DiceCELoss (excluding background, with sigmoid activation and squared prediction enabled, smooth_dr =1e–5). During training, the batch size was configured as 8 images per batch over 200 total epochs, with mixed precision training enabled. In the fine-tuning phase, pre-trained weights were loaded to ensure efficient convergence and adaptation to the segmentation task requirements on both Center A and Center B datasets.

Reliability analysis of segmentation results

The reliability analysis in this study focuses on the results of three core segmentation methods supported by VISTA3D compared against manually annotated gold standards on an external test cohort. The specific sources are as follows: first, automatic segmentation relies on the shared SegResNet backbone and automatic branching to achieve end-to-end segmentation of target structures without additional manual intervention. Second, point-prompt segmentation (interactive segmentation) activates an interactive branch by inputting 3D positive/negative point prompts, flexibly adapting to segmentation demands for new categories not covered by the automatic branch. Third, automatic segmentation + point-prompt segmentation employs a model-specific merging algorithm to use point prompts solely for correcting false positive/false negative regions in the automatic segmentation, avoiding disruption of correctly segmented areas.

To ensure an objective evaluation of the interactive segmentation performance, the 3D point prompts used in the testing

https://dx.doi.org/10.21037/tlcr-2026-1-0023

phase were generated through an automated simulation algorithm rather than manual input. This algorithm identifies the centers of the largest false-negative and false-positive connected components by comparing the automated mask with the manual gold standard, thereby simulating expert correction in a reproducible and unbiased manner.

To prevent data bias from interfering with reliability analysis, all segmentation results must meet the following preprocessing consistency requirements:

(I) Spatial coordinate alignment: map all three segmentation masks and manually segmented gold standards to the coordinate system of the same CT image (based on positioning parameters in DICOM files, ensuring pixel coordinates correspond one-to-one with anatomical locations);

(II) Resolution standardization: resample all images and segmentation masks to a 1 mm × 1 mm × 1 mm isotropic resolution;

(III) Region cropping: retain only segmentation masks within the largest connected region containing the tumor area.

Four widely recognized quantitative metrics in medical image segmentation were selected: Dice similarity coefficient, 95% Hausdorff distance, mean surface distance, and volume similarity. These metrics were used to calculate the consistency between the three segmentation results and the manually segmented gold standard.

Appendix 3

Loss function design

This study employs a composite loss function (L_{total}) to optimize end-to-end DL models. The objective is to simultaneously enhance the accuracy of treatment-specific survival prediction, improve the discriminatory power between different treatment regimens, and ensure temporal consistency in predictions. The total loss function comprises three core components:

$$L_{total}=\lambda_1 L_{survival}+\lambda_2 L_{contrast}+\lambda_3 L_{consistency} \quad [1]$$

(I) Treatment-specific survival prediction loss $L_{survival}$

Objective: accurately predict the PFS risk for patients receiving each radiotherapy regimen.

Mathematical definition: for each radiotherapy regimen, we define a negative log-likelihood loss based on the Cox proportional hazards model:

$$L_{cox}^{k}=-\frac{1}{N_k^{event}}\sum_{i=1}^{N_k}\delta_i^k\left[f^k(X_i)-\log\left(\sum_{j:t_j^k\geq t_i^k}\exp\left(f^k(X_j)\right)\right)\right] \quad [2]$$

Parameter explanation:

- N_k : number of patients actually receiving regimen k;
- N_k^{event} : number of patients observed to experience disease progression under regimen k;
- $\delta_i^k\in\{0,1\}$: censoring indicator; $\delta_i^k=1$ indicates patient i experienced a progression event, while $\delta_i^k=0$ indicates data loss at time t_i^k;
- t_i^k: PFS time for patient i (months);
- $f^k(X_i)$ k: DL model predicted protocol specificity risk score for patient i based on CT images X_i;
- Summation symbol $\sum_{j:t_j^k\geq t_i^k}$ traverses all patients still at risk at time t_i^k.

Conversion to absolute survival probability: to convert the continuous relative risk score $f^k(X_i)$ into the absolute survival probability function $\hat{S}^k(t|X)$ required for downstream loss calculations and clinical assessment, the Breslow estimator was employed to calculate the baseline cumulative hazard $\hat{H}_0^k(t)$:

$$\hat{H}_0^k(t)=\sum_{t_j\leq t}\frac{d_j}{\sum_{i\in R(t_j)}\exp\left(f^k(X_i)\right)} \quad [3]$$

where d_j is the number of events at time t_j, and $R(t_j)$ is the risk set at t_j. The absolute survival probability for a patient with

https://dx.doi.org/10.21037/tlcr-2026-1-0023

imaging features X under regimen k is then calculated as:

$$\hat{S}^k(t|X) = \exp\left(-\hat{H}_0^k(t)\exp\left(f^k(X)\right)\right) \quad [4]$$

Total survival loss:

$$L_{survival}=L_{cox}^{A}+L_{cox}^{B} \quad [5]$$

(II) Treatment comparison loss $L_{contrast}$

Objective: enhance the model's ability to distinguish efficacy differences between two radiotherapy regimens, aligning high-risk score discrepancies with clinically anticipated benefits.

Mathematical definition: triplet margin loss is employed:

$$L_{contrast}=\frac{1}{N_{triplet}}\sum_{(a,p,n)\in T}\max\left(0,\alpha+d_{ap}-d_{an}\right) \quad [6]$$

Parameter explanation:

- $N_{triplet}$: number of valid triplets;
- T: triplet set, where each triplet (a,p,n) comprises:
  - Anchor patient a: serves as the comparison baseline;
  - Positive sample patient p: expected to share similar treatment advantage direction with the anchor patient;
  - Negative sample patients n: expected to exhibit opposite treatment preference direction compared to the anchor patient;
- $\alpha>0$: marginal parameter controlling the minimum distance difference between positive and negative sample pairs (set to $\alpha=0.5$);
- d_{ap} and d_{an}: distance metric for treatment benefit differences, defined as:

$$d_{ij}=\left|\left[f^B(X_i)-f^A(X_i)\right]-\left[f^B(X_j)-f^A(X_j)\right]\right| \quad [7]$$

(III) Time consistency loss Lconsistency

Objective: ensure predicted survival functions exhibit reasonable temporal dynamics consistent with fundamental survival analysis assumptions.

Because the standard baseline hazard estimator inherently yields a step function, continuous analytical derivatives are undefined at event times. Therefore, in our implementation, the first and second derivatives of the survival function were mathematically approximated using finite differences over the discretized evaluation time points $\mathbf{T}$.

Mathematical definition: constraints on the approximated first ($\mathbf{D^1}$) and second ($\mathbf{D^2}$) finite differences of the predicted survival function across discrete time points $\mathbf{t_m} \in \mathbf{T}$:

$$\mathbf{L_{consistency}} = \sum_{k\in\{A,B\}}\sum_{t_m\in T}\left(\lambda_{t1}\cdot\max\left(0,\mathbf{D^1}(t_m)\right)+\lambda_{t2}\cdot\left|\mathbf{D^2}(t_m)\right|\right) \quad [8]$$

where the finite differences for a given time index $\mathbf{m}$ are defined as:

$$\mathbf{D^1}(t_m) = \frac{\hat{S}^k(t_{m+1})-\hat{S}^k(t_m)}{t_{m+1}-t_m} \quad [9]$$

$$\mathbf{D^2}(t_m) = \frac{\mathbf{D^1}(t_m)-\mathbf{D^1}(t_{m-1})}{t_{m+1}-t_{m-1}} \quad [10]$$

Parameter explanation:

- $\hat{S}^k(t|X)$: survival probability of the model-predicted regimen k at time t;
- $T=\{3,6,9,12,18,24,36\}$: discretized time point set (unit: months);

https://dx.doi.org/10.21037/tlcr-2026-1-0023

- $\boldsymbol{\lambda_{t1}}$: monotonicity violation penalty weight, increasing ast increases (emphasizing long-term consistency);
- $\boldsymbol{\lambda_{t2}}$: smoothness penalty weight, preventing sharp fluctuations in survival probability.

Specific constraints:

- Monotonic non-incrementality: survival probability should not increase over time (i.e., $\mathbf{D^1(t_m) \leq 0}$);
- Smoothness constraint: the second finite difference of the survival function should be minimized to avoid unreasonable inflection points.

Final form of total loss function:

$$\mathbf{L_{total} = 1.0 \cdot L_{survival} + 0.3 \cdot L_{contrast} + 0.1 \cdot L_{consistency}} \quad [11]$$

(IV) Optimization algorithm and training strategy

AdamW optimizer combined with weight decay regularization:

$$\theta_{t+1}=\theta_t-\eta\left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon}+\beta\theta_t\right) \quad [12]$$

Parameter settings:

- Initial learning rate $\eta=1\times10^{-4}$;
- Weight decay coefficient $\beta=1\times10^{-4}$;
- First-order moment decay $\beta_1=0.9$;
- Second-order moment decay $\beta_2=0.999$;
- Numerical Stability Constant $\epsilon=1\times10^{-8}$.

Learning rate scheduling: employing cosine annealing strategy:

$$\eta_t=\eta_{min}+\frac{1}{2}(\eta_{max}-\eta_{min})\left(1+\cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right) \quad [13]$$

- $\eta_{min}=1\times10^{-6}$;
- $\eta_{max}=1\times10^{-4}$;
- T_{cur}: current training iteration;
- T_{max}: total training rounds (set to 100 rounds).

Early stopping strategy: monitor AUC on the independent validation set; if no improvement occurs over 10 consecutive epochs, stop training.

(V) Calculation of treatment advantage score

After model training completes, the treatment advantage score for new patients $\mathbf{X}$ is calculated as the mean difference in absolute survival probabilities across the predefined discretized time points $\mathbf{T}$:

$$ASI(X)=\frac{1}{|T|}\sum_{t\in T}\left[\hat{S}^B(t|X)-\hat{S}^A(t|X)\right] \quad [14]$$

Decision rules:

(I) $ASI(X) > 0.5$: recommend hyperfractionated radiotherapy;

(II) $ASI(X) \leq 0.5$: recommend conventional fractionation radiotherapy.

Detailed calculation method for the optimal 0.5 threshold: the decision threshold of 0.5 was not selected arbitrarily, nor does it imply a theoretical requirement for a 50% absolute survival advantage. Instead, 0.5 is the empirically calculated optimal cutoff value ($\mathbf{c_{opt}}$), derived through a rigorous data-driven threshold optimization process on the training cohort using the Maximally Selected Rank Statistics method. The detailed calculation steps are as follows:

(I) Definition of candidate set: we defined a discrete set of candidate thresholds $\mathbf{C} = \{\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_m}\}$ spanning the range of the generated **ASI** scores in the training set;

(II) Subgroup partitioning: for each candidate threshold$\mathbf{c} \in \mathbf{C}$, the training cohort was partitioned into a high-advantage

https://dx.doi.org/10.21037/tlcr-2026-1-0023

subgroup ($\mathbf{ASI(X) > c}$) and a low-advantage subgroup ($\mathbf{ASI(X) \leq c}$);

(III) Statistical evaluation: within the high-advantage subgroup for each threshold $\mathbf{c}$, we computed the log-rank test statistic $\boldsymbol{\chi^2(c)}$ to evaluate the PFS difference between patients who actually received hyperfractionated radiotherapy (BID) and those who received conventional fractionated radiotherapy (QD);

(IV) Optimization objective: the optimal threshold $\mathbf{c_{opt}}$ was defined as the value that maximized the statistical difference (i.e., maximized the prognostic differentiation) between the two treatment regimens within the recommended subgroup:

$$\mathbf{c_{opt} = arg\ \max_{c \in C} \chi^2(c)} \quad [15]$$

(V) Final determination: applying this objective function to our training dataset mathematically yielded $\mathbf{c_{opt} = 0.5}$.

Furthermore, an internal sensitivity analysis was conducted by evaluating adjacent thresholds (e.g., 0.4, 0.6) utilizing the same log-rank methodology. The results confirmed that the calculated threshold of 0.5 provided the most robust and statistically significant patient stratification (maximizing the HR difference for PFS and OS) without the need for manual calibration.

Appendix 4

The detailed methodology of ResNet18

The shared feature extractor is based on the classic ResNet18 convolutional neural network architecture, with all its operations extended from two-dimensional (2D) to 3D to accommodate the processing of medical volumetric data. The input consists of preprocessed image patches of the tumor region of interest. The initial component of the network is a 3D convolutional layer, which employs a relatively large kernel (7×7×7) and a stride of 2 for preliminary feature extraction and down-sampling.

To optimize the model's performance for the specific task of predicting radiotherapy treatment outcomes and to effectively leverage pre-trained weights, this study designed a three-stage training strategy involving two rounds of parameter unfreezing. The core of this strategy is to first train the network layers most directly relevant to the task, followed by gradually unfreezing the deep and then the shallow layers of the shared feature extractor in two distinct phases. This allows the model to progressively adapt from generic features to the specific medical imaging prediction task of this study.

Stage 1: prediction branch training (all backbone layers frozen)

The objective of this stage is to initialize the two treatment-specific prediction branches. We employ a 3D ResNet18 model, pre-trained on the large-scale natural image dataset (ImageNet), as the shared feature extractor and completely freeze all its parameters. The model input consists of preprocessed, tumor-centered 3D CT image patches. During this stage, only the parameters of the fully connected layers within the two prediction branches are trained and updated. Training utilizes the fundamental component of the composite loss function—the treatment-specific conditional survival loss ($L_{survival}$). The AdamW optimizer is employed with an initial learning rate set to a relatively low value of 5×10^{-5} to prevent excessively large gradients during the random initialization of the branch parameters. This stage continues for 20 epochs, aiming to rapidly establish an initial mapping relationship from features to risk scores based on fixed and relatively generic image features, thereby laying the foundation for subsequent fine-tuning.

Stage 2: fine-tuning of deep semantic features (first unfreezing: the last two residual stages)

After the prediction branches have stabilized initially, we perform the first parameter unfreezing. The target of this unfreezing is the last two residual stages (stages 3 and 4) within the shared ResNet18 backbone network. These two stages contain deeper network layers responsible for encoding high-level, semantically rich features, which are considered to be closely

https://dx.doi.org/10.21037/tlcr-2026-1-0023

associated with the biological behavior of tumors and their response to treatment. After unfreezing, the weights of these deep convolutional layers will be fine-tuned based on the survival prediction task of this study. Meanwhile, the first two residual stages of the network (stages 1 and 2), along with all batch normalization layer parameters, remain frozen. These initial stages primarily extract low-level generic features (such as edges and corners), which remain relatively stable in medical images. Preserving their pre-trained weights helps maintain the model's generalization capability and prevents overfitting. The training cycle for this stage is increased to 30 iterations, with the learning rate adjusted to 1×10^{-5}.

Phase three: end-to-end fine-tuning (second thaw: remaining trainable layers)

To achieve optimal synergy between feature extraction and treatment prediction, we perform a second parameter thawing. This thawing unfreezes all remaining parameters previously frozen within the ResNet18 backbone network, encompassing both the first and second stages. At this point, all trainable parameters across the entire model—including every layer of the feature extractor and both prediction branches—are released for end-to-end joint fine-tuning. This stage enables the lower-level feature extractor to adaptively adjust based on the upstream survival prediction task, ensuring the entire feature learning process is highly aligned with the ultimate goal of predicting treatment efficacy. This training phase employs 20–40 epochs (using an early stopping strategy), utilizing a lower fixed learning rate of 5×10^{-6}. It activates the full composite triplet survival contrastive loss function, which integrates survival loss, contrastive loss, and temporal consistency regularization to comprehensively optimize the model's predictive accuracy, discriminative power, and clinical plausibility.

https://dx.doi.org/10.21037/tlcr-2026-1-0023

Table S1 Baseline characteristics of patients receiving BID *vs.* QD radiotherapy, overall and stratified by ASI score

Variables	Overall (n=446)			ASI >0.5 (n=221)			ASI ≤0.5 (n=225)		
	BID (n=223)	QD (n=223)	P	BID (n=112)	QD (n=109)	P	BID (n=111)	QD (n=114)	P
Gender			0.892			0.764			0.831
Female	61 (27.35)	62 (27.80)		30 (26.79)	31 (28.44)		31 (27.93)	31 (27.19)	
Male	162 (72.65)	161 (72.20)		82 (73.21)	78 (71.56)		80 (72.07)	83 (72.81)	
Age (years)			0.674			0.582			0.719
<60	112 (50.22)	108 (48.43)		57 (50.89)	52 (47.71)		55 (49.55)	56 (49.12)	
≥60	111 (49.78)	115 (51.57)		55 (49.11)	57 (52.29)		56 (50.45)	58 (50.88)	
Smoking			0.581			0.493			0.627
No	117 (52.47)	113 (50.67)		59 (52.68)	54 (49.54)		58 (52.25)	59 (51.75)	
Yes	106 (47.53)	110 (49.33)		53 (47.32)	55 (50.46)		53 (47.75)	55 (48.25)	
ECOG			0.735			0.618			0.804
0	114 (51.12)	110 (49.33)		58 (51.79)	54 (49.54)		56 (50.45)	56 (49.12)	
1	109 (48.88)	113 (50.67)		54 (48.21)	55 (50.46)		55 (49.55)	58 (50.88)	
Location			0.812			0.705			0.893
Left	105 (47.09)	102 (45.74)		53 (47.32)	49 (44.95)		52 (46.85)	53 (46.49)	
Right	118 (52.91)	121 (54.26)		59 (52.68)	60 (55.05)		59 (53.15)	61 (53.51)	
T stage			0.384			0.421			0.536
T1	47 (21.08)	44 (19.73)		24 (21.43)	22 (20.18)		23 (20.72)	22 (19.30)	
T2	78 (34.98)	83 (37.22)		39 (34.82)	41 (37.61)		39 (35.14)	42 (36.84)	
T3	41 (18.39)	39 (17.49)		21 (18.75)	19 (17.43)		20 (18.02)	20 (17.54)	
T4	57 (25.56)	57 (25.56)		28 (25.00)	27 (24.77)		29 (26.13)	30 (26.32)	
N stage			0.629			0.548			0.712
N0–1	45 (20.18)	43 (19.28)		23 (20.54)	21 (19.27)		22 (19.82)	22 (19.30)	
N2–3	178 (79.82)	180 (80.72)		89 (79.46)	88 (80.73)		89 (80.18)	92 (80.70)	
CCRT			0.847			0.763			0.891
No	68 (30.49)	68 (30.49)		34 (30.36)	32 (29.36)		34 (30.63)	36 (31.58)	
Yes	155 (69.51)	155 (69.51)		78 (69.64)	77 (70.64)		77 (69.37)	78 (68.42)	
Chemotherapy			0.208			0.315			0.427
EC	109 (48.88)	101 (45.29)		55 (49.11)	49 (44.95)		54 (48.65)	52 (45.61)	
EP	114 (51.12)	122 (54.71)		57 (50.89)	60 (55.05)		57 (51.35)	62 (54.39)	
PCI			0.419			0.502			0.613
No	108 (48.43)	102 (45.74)		55 (49.11)	50 (45.87)		53 (47.75)	52 (45.61)	
Yes	115 (51.57)	121 (54.26)		57 (50.89)	59 (54.13)		58 (52.25)	62 (54.39)	

Data are presented as n (%). BID, twice daily; CCRT, concurrent chemoradiotherapy; EC, etoposide and carboplatin; ECOG, Eastern Cooperative Oncology Group; EP, etoposide and cisplatin; N, node; PCI, percutaneous coronary intervention; QD, once daily; T, tumor.

https://dx.doi.org/10.21037/tlcr-2026-1-0023

Table S2 Multivariable Cox regression analysis for PFS and OS

Variables	PFS		OS	
	HR (95% CI)	P	HR (95% CI)	P
Treatment (BID *vs.* QD)	0.89 (0.71–1.12)	0.312	0.92 (0.74–1.15)	0.458
ASI score (continuous)	0.76 (0.62–0.93)	0.008*	0.71 (0.58–0.87)	0.001*
Treatment × ASI interaction	0.63 (0.45–0.88)	0.007*	0.59 (0.42–0.83)	0.002*
Age (≥60 *vs.* <60 years)	1.18 (0.96–1.45)	0.112	1.24 (1.01–1.52)	0.041*
Gender (male *vs.* female)	1.05 (0.84–1.31)	0.674	1.08 (0.86–1.36)	0.512
Smoking (yes *vs.* no)	1.12 (0.91–1.38)	0.283	1.15 (0.93–1.42)	0.196
ECOG (1 *vs.* 0)	1.09 (0.95–1.58)	0.422	1.14 (0.89–1.45)	0.187
Location (right *vs.* left)	1.04 (0.85–1.27)	0.712	1.02 (0.83–1.25)	0.847
T stage				
T2 *vs.* T1	1.21 (0.91–1.61)	0.189	1.18 (0.88–1.58)	0.264
T3 *vs.* T1	1.45 (1.05–2.00)	0.024*	1.42 (1.02–1.98)	0.037*
T4 *vs.* T1	1.68 (1.23–2.29)	0.001*	1.73 (1.26–2.38)	<0.001*
N stage (N2–3 *vs.* N0–1)	1.02 (0.88–1.26)	0.667	1.18 (0.78–1.32)	0.698
Chemotherapy (EP *vs.* EC)	0.91 (0.74–1.12)	0.367	0.89 (0.72–1.10)	0.284
PCI (yes *vs.* no)	0.82 (0.67–1.01)	0.062	0.79 (0.64–1.06)	0.084

*, P<0.05. BID, twice daily; CI, confidence interval; EC, etoposide and carboplatin; ECOG, Eastern Cooperative Oncology Group; EP, etoposide and cisplatin; HR, hazard ratio; N, node; OS, overall survival; PCI, percutaneous coronary intervention; PFS, progression-free survival; QD, once daily; T, tumor.

https://dx.doi.org/10.21037/tlcr-2026-1-0023