Inclusion and exclusion criteria

The following inclusion criteria were applied: (I) patients aged ≥18 years; (II) primary or recurrent soft tissue sarcoma (STS) of the extremities or trunk (excluding unplanned tumor resection from other hospitals that require extended resection), with two of the following characteristics: high grade, ≥5 cm or deep-seated tumor, after multidisciplinary team discussion, requiring radiotherapy; (III) Eastern Cooperative Oncology Group (ECOG) performance status 0–3; (IV) pathologically confirmed STS; (V) female patients of childbearing age or male patients with female partners of childbearing age must use effective contraception throughout the treatment period and for 6 months after the last dose of medication; (VI) magnetic resonance imaging (MRI) was performed before and after neoadjuvant radiotherapy (neo-RT), and the image quality was good.

The exclusion criteria were as follows: (I) patients who have undergone unplanned tumor resection from other hospitals and have no residual macroscopic tumor; (II) patients with other comorbidities that preclude the use of anlotinib; (III) new primary malignant tumors within the past 5 years (except for cervical carcinoma *in situ* or early-stage basal cell carcinoma of the skin); (IV) STSs that can be cured by simple excision; (V) history of radiotherapy to the same site radiotherapy (RT); (VI) presence of distant metastasis; (VII) planned to undergo neoadjuvant chemotherapy, use of chemotherapy or other drug treatments (including anlotinib) within the past 4 weeks; (VIII) RT was not completed as planned; (IX) MRI images before and/or after RT were missing or could not be compared.

Treatment and clinical efficacy assessment

All patients included in this study received neoadjuvant RT with a total dose of 50 Gy in 25 fractions, with concurrent and sequential anlotinib (12 mg/day for 2 weeks every 3 weeks, with a total of 3 cycles) or apatinib (500 mg/day, 5 days per week, started from 2 weeks before RT to 4 weeks after RT) treatment, followed by planned limb-conserving wide resection (1–2 weeks apart from anlotinib or apatinib and 4–8 weeks apart from RT).

All postoperative resected specimens were reviewed and evaluated by a senior pathologist with 19 years of experience in STS diagnosis who was blinded to the clinical and MRI data. Good-response and poor-response group was defined as <50% and ≥50% of residual tumor cells as per postoperative pathology. Clinical characteristics, including age, sex, tumor location, tumor size, T stage and grade were obtained for all patients. Grading was evaluated according to the French Federation of Cancer Centers Sarcoma Group (FNCLCC) grading system. T stage was evaluated according to the 8th edition of the AJCC Cancer Staging Manual. The RECIST 1.1 criteria were used to compare the tumor size changes. Informed consent was obtained before treatment.

Protocols for the MRI IVIM examination

For IVIM imaging, a single-shot echo-planar imaging (SS-EPI) sequence was utilized, acquiring images at 14 distinct b-values (10, 20, 40, 60, 80, 100, 150, 200, 400, 800, 1,000, 1,200, 1,500 s/mm^2) in three orthogonal directions. The detailed imaging parameters for the IVIM study were as follows: repetition time/echo time (TR/TE), auto/minimum; section thickness, 4–5 mm; gap, 1 mm; field of view (FOV), 25–35 cm; matrix, 128×128; number of excitations (NEX), 1–6; bandwidth, 250 Hz/pixel. IVIM imaging acquisition were performed prior to contrast agent administration.

Bi-exponential equation

The IVIM model and its parameters were fitted according to the following bi-exponential equation:

$$S(b)/S_0 = f \cdot \exp(-bD^*) + (1-f) \cdot \exp(-bD)$$

where S(b) is the signal intensity at a specific b value, S_0 is the signal intensity at b=0 s/mm^2, f represents perfusion fraction. D^* is the perfusion-related diffusion coefficient, and D represents the diffusion of the non-perfusing fraction.

Optimization of cluster numbers

The voxels were grouped using the K-means clustering algorithm, based on the squared Euclidean distances between voxel intensities. All voxels were allocated into one of the clusters and were displayed as spatial habitats in the original image space. As the optimum number of clusters in a dataset is an important issue in K-means clustering, 2, 3, 4, and 5 clusters were initially set. The algorithm iteratively adjusted cluster centers until convergence was achieved, defined by no changes in cluster centers, or until at most 500 iterations were reached. We used the silhouette score to

https://dx.doi.org/10.21037/tcr-2025-aw-2201

evaluate the optimal number of clusters, as it measures how well each object lies within its cluster, defines the average distance (cohesion) between an object and other samples in the same cluster, and the average distance (separability) between an object and all samples in the other clusters. Our results showed that the silhouette score reached an inflection point at 3 clusters (*Figure S1*), so we finally chose 3 clusters. The whole tumor can be divided into two regions based on the levels of *D*, and the group with lower *D* can be further divided into two categories based on the levels of *D**.

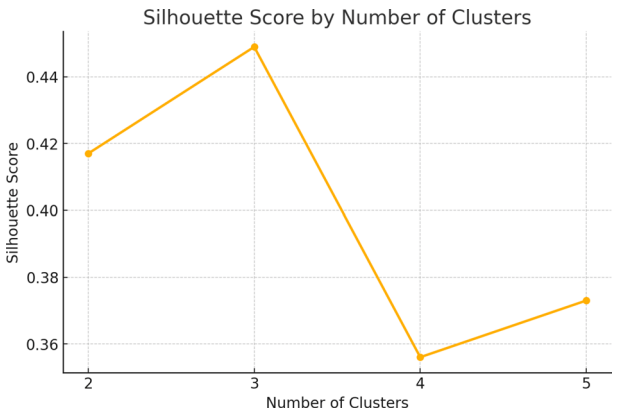

Figure S1 Silhouette score by number of clusters.

References

43. Wardelmann E, Haas RL, Bovée JV, et al. Evaluation of response after neoadjuvant treatment in soft tissue sarcomas; the European Organization for Research and Treatment of Cancer-Soft Tissue and Bone Sarcoma Group (EORTC-STBSG) recommendations for pathological examination and reporting. Eur J Cancer 2016;53:84-95.

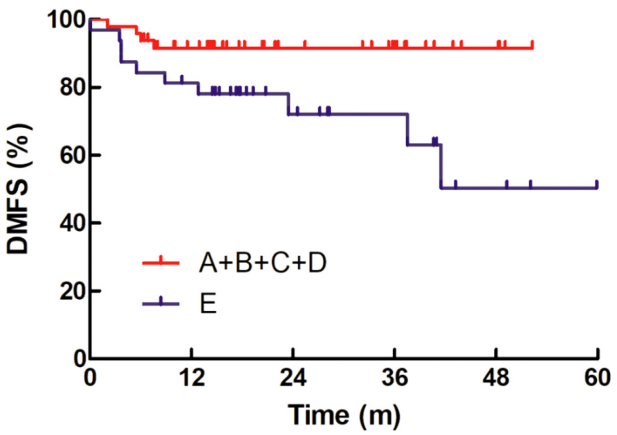

Figure S2 DMFS was assessed for 80 soft tissue sarcoma patients who were categorized into either group A + B + C + D and group E. Group A + B + C + D had better DMFS than that of grade E. A–E grades were classified according to EORTC-STBSG recommendations for pathological examination and reporting (43). In detail, the categories are defined as follows: A, no stainable tumor cells; B, single stainable tumor cells or small clusters (overall below 1% of the whole specimen); C, ≥1% and <10% stainable tumor cells; D, ≥10% and <50% stainable tumor cells; E, ≥50% stainable tumor cells. DMFS, distant metastasis free survival; EORTC-STBSG, European Organization for Research and Treatment of Cancer-Soft Tissue and Bone Sarcoma Group.

https://dx.doi.org/10.21037/tcr-2025-aw-2201

Table S1 Comparison of the diagnostic performance of all models using the DeLong method

Model 1	Model 2	AUC of model 1	AUC of model 2	P value
$D^*_{\text{Cluster 2 Mean (Pre)}}$	RECIST 1.1	0.856	0.567	0.03*
$f_{\text{Cluster 1 Mean (Post)}}$	RECIST 1.1	0.865	0.567	0.04*
Combined model	RECIST 1.1	0.904	0.567	0.02*
$D_{\text{Cluster 1 Mean (Post)}}$	RECIST 1.1	0.837	0.567	0.09
$D_{\text{whole Mean (Post)}}$	RECIST 1.1	0.808	0.567	0.11
$f_{\text{Cluster 3 Mean (Post)}}$	Combined model	0.779	0.904	0.17
$f_{\text{Cluster 3 Mean (Post)}}$	RECIST 1.1	0.779	0.567	0.28
$D_{\text{Cluster 1 Mean (Post)}}$	Combined model	0.837	0.904	0.30
$D_{\text{whole Mean (Post)}}$	Combined model	0.808	0.904	0.29
$D^*_{\text{Cluster 2 Mean (Pre)}}$	$f_{\text{Cluster 3 Mean (Post)}}$	0.856	0.779	0.55
$D^*_{\text{Cluster 2 Mean (Pre)}}$	$D_{\text{whole Mean (Post)}}$	0.856	0.808	0.69
$D^*_{\text{Cluster 2 Mean (Pre)}}$	Combined model	0.856	0.904	0.50
$f_{\text{Cluster 1 Mean (Post)}}$	$f_{\text{Cluster 3 Mean (Post)}}$	0.865	0.779	0.49
$f_{\text{Cluster 1 Mean (Post)}}$	$D_{\text{whole Mean (Post)}}$	0.865	0.808	0.61
$f_{\text{Cluster 1 Mean (Post)}}$	Combined model	0.865	0.904	0.63
$f_{\text{Cluster 3 Mean (Post)}}$	$D_{\text{Cluster 1 Mean (Post)}}$	0.779	0.837	0.61
$D_{\text{Cluster 1 Mean (Post)}}$	$D_{\text{whole Mean (Post)}}$	0.837	0.808	0.71
$f_{\text{Cluster 1 Mean (Post)}}$	$D_{\text{Cluster 1 Mean (Post)}}$	0.865	0.837	0.76
$f_{\text{Cluster 3 Mean (Post)}}$	$D_{\text{whole Mean (Post)}}$	0.779	0.808	0.81
$D^*_{\text{Cluster 2 Mean (Pre)}}$	$D_{\text{Cluster 1 Mean (Post)}}$	0.856	0.837	0.86
$D^*_{\text{Cluster 2 Mean (Pre)}}$	$f_{\text{Cluster 1 Mean (Post)}}$	0.856	0.865	0.94

*, P<0.05.

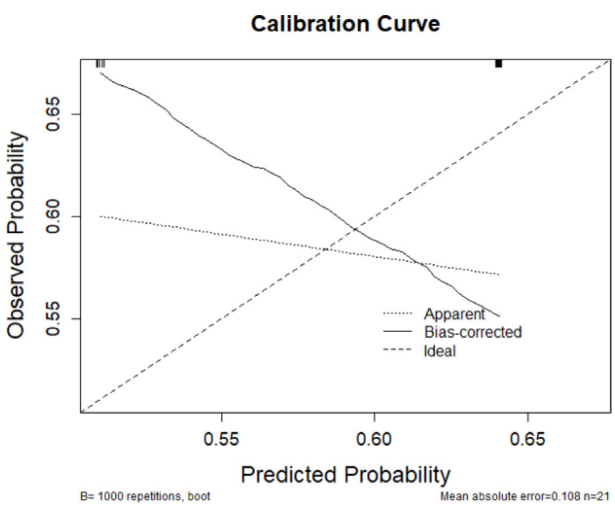

Figure S3 Calibration curve for RECIST 1.1.

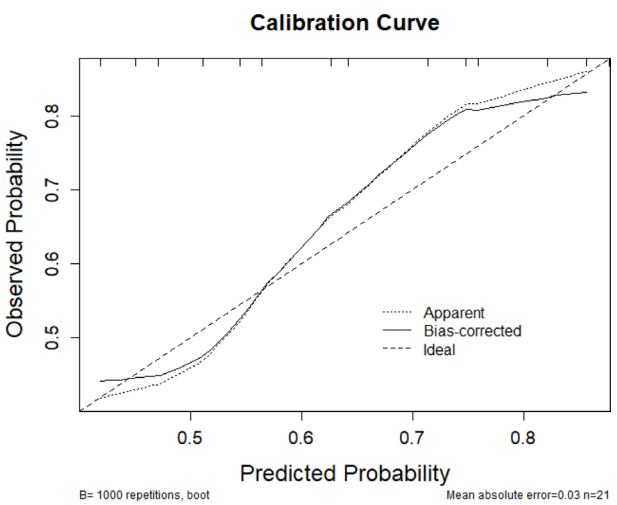

Figure S4 Calibration curve for $f_{\text{Cluster 3 Mean (Post)}}$.

https://dx.doi.org/10.21037/tcr-2025-aw-2201

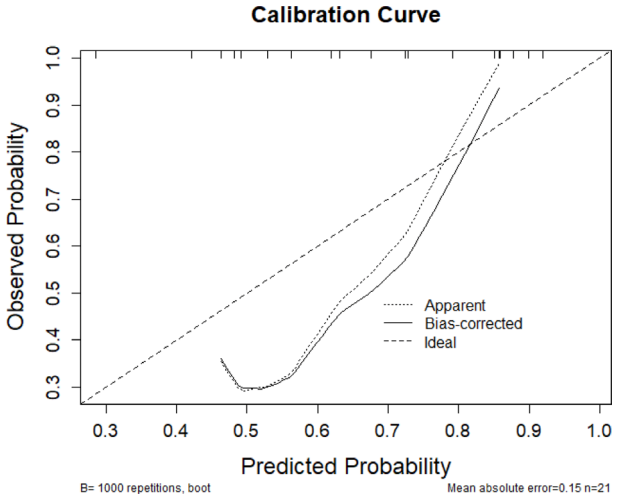

Figure S5 Calibration curve for $D_{\text{whole Mean (Post)}}$.

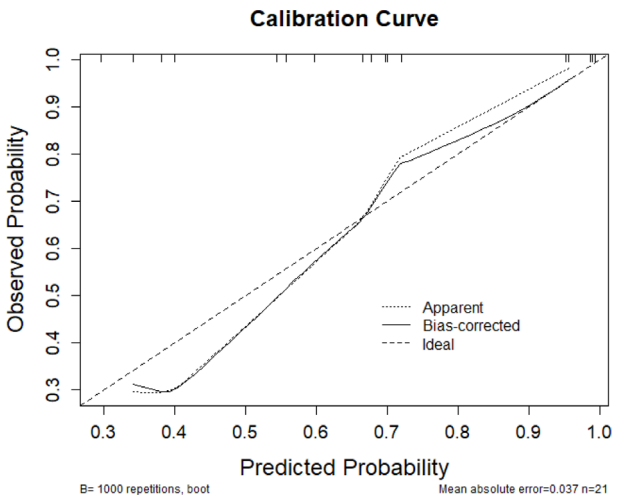

Figure S6 Calibration curve for $D_{\text{Cluster 1 Mean (Post)}}$.

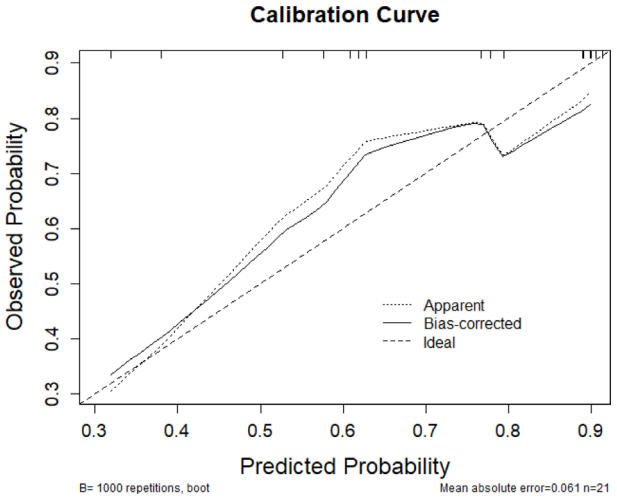

Figure S7 Calibration curve for $D^{*}_{\text{Cluster 2 Mean (Pre)}}$.

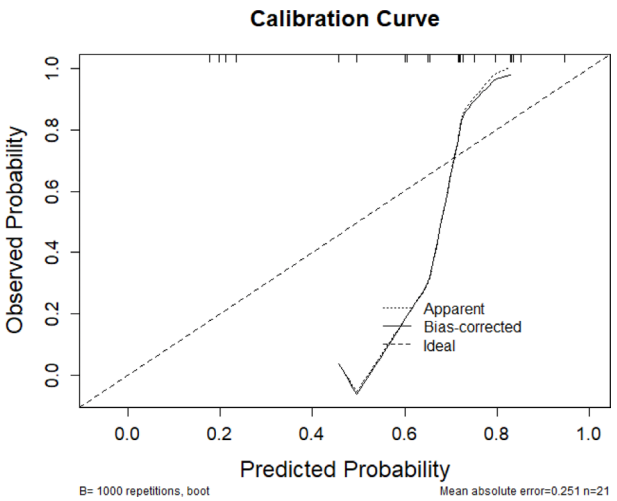

Figure S8 Calibration curve for $f_{\text{Cluster 1 Mean (Post)}}$.

https://dx.doi.org/10.21037/tcr-2025-aw-2201

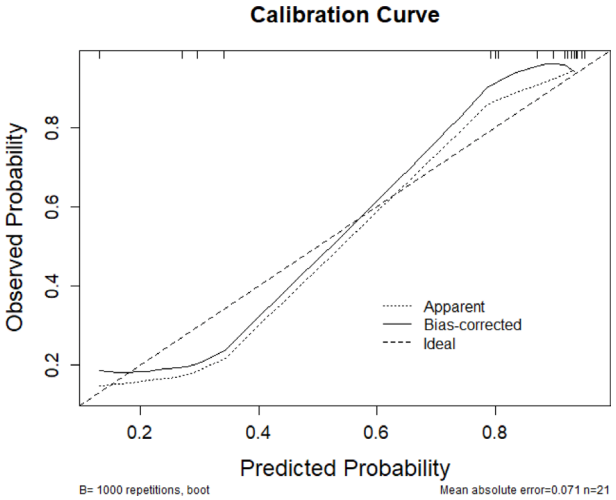

Figure S9 Calibration curve for combined model.

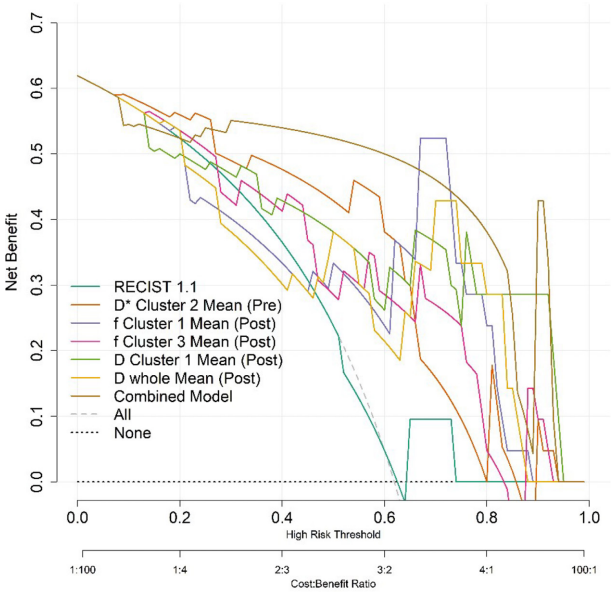

Figure S10 DCA of all seven models. The net benefit was determined by deducting the rate of false-positive cases from the rate of true-positive cases, weighted by the relative harm of a false-positive and -negative result. The two extreme strategies, "treat all" and "treat none", were used as benchmarks. A decision model was considered clinically advantageous if its decision curve outperformed both reference strategies in terms of the net benefit. DCA, decision curve analysis.

https://dx.doi.org/10.21037/tcr-2025-aw-2201