Am J Cancer Res 2026;16(4):1327-1340
www.ajcr.us /ISSN:2156-6976/ajcr0171281

Original Article

Development and temporal validation of a machine learning-based model to predict postoperative recurrence and guide adjuvant radiotherapy in patients with soft tissue sarcoma: a retrospective cohort study

Zhenguo Zhao[1*], Xinyu Li[1*], Xinfeng Wang[2*], Xu Liu[3*], Jin Yuan[4], Xiaoyang Li[1], Luqiang Wang[1], Xinxin Zhang[1], Ting Liu[1], Shengji Yu[1]

[1]Department of Orthopedics, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; [2]Department of General Surgery, Beijing Friendship Hospital, Capital Medical University, Beijing 101100, China; [3]Department of Pancreatic and Gastric Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; [4]Department of Orthopedics, Xuanwu Hospital, Capital Medical University, No. 45 Changchun Street, Xicheng District, Beijing 100053, China. [*]Equal contributors.

Received January 6, 2026; Accepted April 7, 2026; Epub April 15, 2026; Published April 30, 2026

Abstract: Background: Soft tissue sarcoma (STS) is highly heterogeneous and has a high risk of recurrence so that the accurate prognosis of postoperative recurrence and the value of radiotherapy are critically important. We developed a machine learning model to predict postoperative recurrence in soft tissue sarcoma patients, providing a data-driven tool to optimize adjuvant radiotherapy decisions. Methods: We retrospectively analyzed 642 STS patients who underwent radical surgery at the China National Cancer Center from 2010 to 2025. In order to determine the essential predictors and build strong models, we used a machine learning approach based on wrapper methods. The performance of models was rigorously evaluated using temporal validation, where the concordance index (C-index), time-dependent receiver operating characteristic (ROC), and decision curve analysis (DCA) were used to identify the most effective predictive architecture. Results: There are a total of 11 feature subsets are identified, which are combined with 11 machine learning algorithms in a combinatorial manner, resulting in 121 predictive models. Among these models, the Cox proportional hazards model combined with Random Survival Forests (COXPH+RSF, CRM) demonstrates the best predictive performance. The C-index for CRM is 0.923 (95% CI 0.878-0.935) in the training cohort, 0.867 (95% CI 0.850-0.875) in the cross-validated training cohort, and 0.807 (95% CI 0.765-0.819) in the temporal validation cohort. Time-dependent calibration curves, time-dependent ROC curves and DCA evaluation confirms that the CRM achieves high predictive precision and clinical utility. We also release an open-access online platform to host our model and to support its practical application. And staging system based on the CRM provides a new clinical reference for determining postoperative adjuvant radiotherapy strategies in this patient cohort. Conclusions: The CRM demonstrates the best predictive capacity concerning recurrence after surgery in STS patients, which could be of immense potential to assist clinicians in assessing disease severity, guiding patient follow-up, and informing adjuvant treatment strategies.

Keywords: Soft tissue sarcoma, cancer recurrence, surgery, machine learning-based model, prognostic model

Introduction

Soft tissue sarcoma (STS) is a rare and heterogeneous malignancy, accounting for about 1 percent of adult and 15 percent of pediatric cancers [1]. There has been more than 50 histological subtypes identified to date, which contributes to significant heterogeneity in prognosis as well as treatment response [2-4]. The multidisciplinary treatment strategies have significantly improved local control rates, but the risk of recurrence and long-term disease control remains challenging in a subset of patients [5, 6]. For example, studies have shown that for

extremity STS, adjuvant radiotherapy can significantly improve local control compared with wide resection alone. In contrast to extremity sarcomas, retroperitoneal soft tissue sarcomas still have a high rate of local recurrence even after complete resection [7]. Additionally, it is not known yet what the impact of postoperative radiotherapy is on long-term overall survival, and whether it provides meaningful advantages over preoperative radiotherapy remains controversial [8].

Current clinical practice relies on multidisciplinary team to develop and individualized treatment plan based on each patient's specific circumstances and tries to find a balance between therapeutic efficacy and potential negative effects. Although the fact that traditional prognostic factors such as tumor grade, size, depth, resection margin status, and metastatic spread are well recognized, their predictive ability in clinical practice remains limited [9]. In particular, existing prognostic models lack reliability in predicting postoperative recurrence risk and the efficacy of adjuvant radiotherapy in patients with STS.

In recent years, machine learning has demonstrated great potential in acquiring intricate clinical information and nonlinear interactions, and its applications in the field of oncology are increasingly emerging [1, 10, 11]. Machine learning has already been used in various cancers to predict postoperative complications, survival outcomes, and recurrence risks, with potential advantages over traditional methods [12-15]. While current research utilizes machine learning and radiomics to predict overall survival and neoadjuvant therapy outcomes in STS patients [1, 16], individualized prediction models in terms of postoperative recurrence risk remain to be further developed.

Our research develops and validates a machine learning-based model to predict recurrence in STS patients, with an addition of a web-based risk calculator to assess the clinical utility of postoperative adjuvant radiotherapy. Consequently, such a model can be used as a pragmatic decision-support tool for personalizing adjuvant radiotherapy in routine practice.

Material and methods

This study was conducted in accordance with the Prediction model Risk Of Bias Assessment Tool (PROBAST) standards and a checklist for useful clinical prediction tools proposed by Florian Markowetz, with reporting guided by Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [17-19]. The complete research process of this study is shown in **Figure 1**.

Study population

The single center retrospective cohort study enrolled 897 patients with STS who underwent radical surgery at the National Cancer Center between 2010 and 2025. Patients were divided into two groups: the training cohort (2010-2020, n=365) and the temporal validation cohort (2021-2025, n=277). Each patient gave written informed consent before surgical intervention and subsequent follow-up. All procedures that involved human participants were performed in accordance with the 1975 Declaration of Helsinki and its amendments, and ethical approval was obtained by the Hospital Ethics Committee of the National Cancer Center (No. NCC2020C).

Patients with pathologically confirmed STS who underwent radical resection were included. They were excluded if they had any of the following: (1) Died during perioperative period; (2) Lost to follow-up; (3) Insufficient clinical data. Out of the initial screening of 897 patients with soft tissue sarcoma, 255 were not eligible according to the inclusion criteria and were therefore excluded. Ultimately, 642 patients were included in the analysis.

Data collection

The surgeries were executed by senior surgeons with specialization in oncological resection. The data were retrospectively collected by a multidisciplinary team of soft tissue sarcoma specialists and clinicians. All variables are commonly used in clinical practice. The sex and age of the patients can be derived from their identification documents. The chemotherapy sensitive subtype, depth, grade, margin, N stage, pathology subtype, and size can be determined from postoperative pathology reports.

Prior to collecting the medical data, all staff underwent standardized training on the data extraction forms to preserve data consistency.

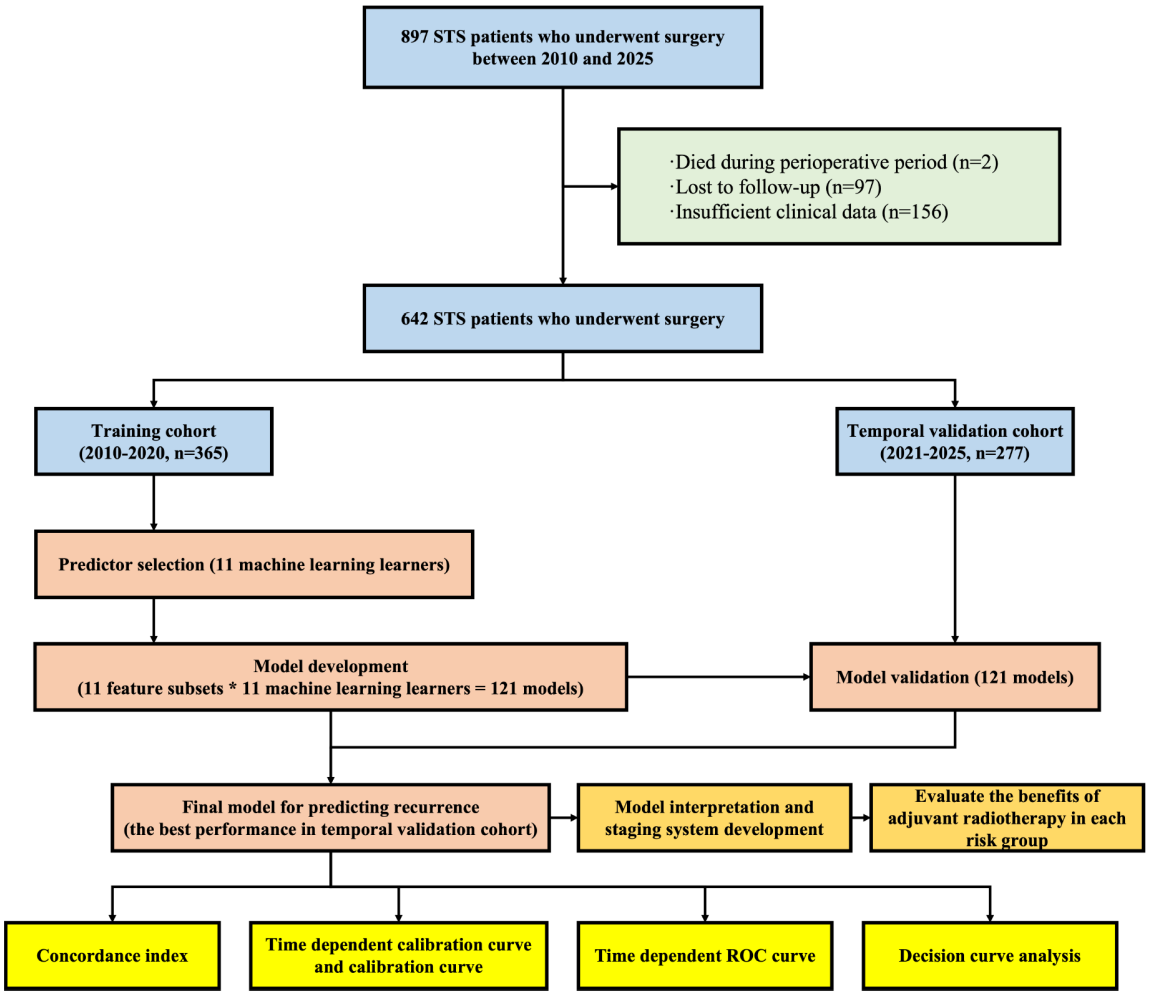


Figure 1. Flow diagram of the study.

After reviewing eligible medical records, outcomes were assessed by an independent panel that was blinded to the model predictions and input variables. This is to ensure the recurrence verification remains strictly based on clinical data of the participants, thereby eliminating potential bias or influence from the model's predictive outcomes.

Predictor selection

In the training cohort, variables with over 35 percent missing data were excluded from the analysis. A total of 14 variables were included in the training set: sex, age, pathological subtype, chemotherapy sensitive subtype, tumor location, depth, size, T stage, N stage, histological grade (FNCLCC), AJCC stage, surgical margin status, postoperative chemotherapy, and postoperative radiotherapy. These variables were entered into 11 machine learning models to choose the predictive factors.

We applied a range of established algorithms that can accommodate both continuous and categorical variables, including Boosted Generalized Additive Model (GAMB), Boosted Generalized Linear Model (GLMB), Survival Tree (ST), Conditional Inference Tree (CIT), Random Survival Forest (RSF), Conditional Random Forest (CRF), Accelerated Oblique Random Survival Forest (AORSF), Penalized Regression (PR), Cox Proportional Hazards Model (COXPH), Kaplan-Meier Estimator (KM), Normalized Absolute Error (NAE). All these algorithms are all sourced from the "mlr3proba" R package [20].

There are three major feature selection methods: filter, wrapper, and embedded. The main purpose of feature selection is to assess each

feature's importance within the full set and identify those most relevant for predicting the target. The present study employed wrapper methods (WM), which evaluate feature usefulness based on classifier performance [21]. The predictor selection using the WM involves four key steps: (1) Implementing sequential forward selection to iteratively build feature subsets; (2) Assessing each subset's performance by calculating the concordance index (C-index) through 10-fold cross-validation; (3) Repeat the above process for all feature subsets; (4) Select and display the feature subset with the highest C-index as the outcome of the WM on that algorithm. The algorithm outputs a specific combination of clinical features as the final predictors.

Machine learning model development and validation

To explore the optimal predictive model, a systematic framework was implemented by cross-pairing 11 machine learning algorithms with 11 independently derived feature subsets. This combinatorial approach ensured a comprehensive evaluation of various algorithm-feature integrations during the model development phase. And these models were later validated in a validation cohort.

The C-index is widely applied in medical research as a statistical measure to evaluate the predictive performance of survival analysis models. It measures how well the alignment predictions correspond to actual results, which range from 0 to 1. Overall, a higher C-index indicates better predictive performance of the regression model. The model with the highest average C-index on validation data was identified as the one for further study.

The predictive performance of the model was measured in a number of ways: calibration curves were established to evaluate the concordance between predicted and actual non-recurrence rates at 1, 3, and 5 years. Meanwhile, time-dependent variants reflected calibration stability over the entire study period. The area under the receiver operating characteristic (ROC) curves (AUC) was used to assess predictive discrimination abilities of the models. Decision curve analysis (DCA) was conducted to evaluate the potential clinical benefits to patients across the same time horizons. Finally, individual risk scores were derived for both the training and validation cohorts using the machine learning model.

We quantified the temporal dynamics of feature influence through a time-dependent importance analysis. The value of each predictor was evaluated by computing the model's Brier score loss after permuting feature values. To enhance reliability, this evaluation was incorporated into a 10-fold cross-validation resampling strategy. This approach allowed determining which variables are of importance to model predictions that vary over time and provide essential evidence for time-sensitive clinical decision making.

Web risk calculator and staging system

Based on the ST, optimal risk score cut-offs were identified to classify patients into high-risk and low-risk groups. The final model was integrated into an interactive web-based application using the R package "Shiny". This web-based application allows clinicians to enter patient data to calculate recurrence risk in STS patients which will be a useful tool to use by health practitioners when making early treatment plans [22].

Statistical analysis

The normality of the data was tested by the Kolmogorov-Smirnov test. The continuous variables that have normal distribution were described as mean ± standard deviation and analyzed by the t-test; otherwise, they were reported as median (interquartile range) and analyzed by the Mann-Whitney U test. In case of categorical variables, the data were presented as numbers and frequencies, applying either the Chi-square test or Fisher's exact test as appropriate. The cumulative risk curves were used to visualize and evaluate the differences of the recurrence rates between the two risk strata (high-risk and low-risk). All the statistical tests were two-sided, with a *p*-value threshold of 0.05 for statistical significance. All figure illustrations and statistical analyses were done in R version 4.4.1.

Results

Patient characteristics

A total of 897 soft tissue sarcoma patients were initially screened and 255 of them were

Table 1. Baseline characteristics of the patients

	Training cohort N=365	Validation cohort N=277	P
Age	51.0 [38.0; 63.0]	58.0 [45.0; 69.0]	<0.001
Sex			0.797
Female	163 (44.7%)	120 (43.3%)	
Male	202 (55.3%)	157 (56.7%)	
Pathology Subtype			<0.001
Liposarcoma	65 (17.8%)	142 (51.3%)	
Undifferentiated Pleomorphic Sarcoma	132 (36.2%)	34 (12.3%)	
Synovial Sarcoma	96 (26.3%)	33 (11.9%)	
Fibrosarcoma	47 (12.9%)	39 (14.1%)	
Other	25 (6.85%)	29 (10.5%)	
Chemotherapy sensitive subtype			<0.001
No	253 (69.3%)	230 (83.0%)	
Yes	112 (30.7%)	47 (17.0%)	
Site			0.613
Upper limbs	58 (15.9%)	40 (14.4%)	
Lower limbs	182 (49.9%)	149 (53.8%)	
Trunk or Pelvis	125 (34.2%)	88 (31.8%)	
Depth			0.905
Superficial	175 (47.9%)	135 (48.7%)	
Deep	190 (52.1%)	142 (51.3%)	
Size	6.00 [3.50; 10.0]	8.00 [4.10; 13.0]	0.001
T			<0.001
T1	137 (37.5%)	79 (28.5%)	
T2	131 (35.9%)	80 (28.9%)	
T3	41 (11.2%)	67 (24.2%)	
T4	56 (15.3%)	51 (18.4%)	
N			0.802
N0	352 (96.4%)	269 (97.1%)	
N1	13 (3.56%)	8 (2.89%)	
Grade			<0.001
G1	208 (57.0%)	120 (43.3%)	
G2	104 (28.5%)	66 (23.8%)	
G3	53 (14.5%)	91 (32.9%)	
AJCC Stage			0.002
IA	74 (20.3%)	34 (12.3%)	
IB	129 (35.3%)	86 (31.0%)	
II	59 (16.2%)	45 (16.2%)	
IIIA	49 (13.4%)	44 (15.9%)	
IIIB	41 (11.2%)	60 (21.7%)	
IV	13 (3.56%)	8 (2.89%)	
Margin			0.029
Negative	301 (82.5%)	208 (75.1%)	
Positive	64 (17.5%)	69 (24.9%)	
Radiotherapy			0.197
No	203 (55.6%)	169 (61.0%)	
Yes	162 (44.4%)	108 (39.0%)	

Chemotherapy			0.007
No	271 (74.2%)	231 (83.4%)	
Yes	94 (25.8%)	46 (16.6%)	
Recurrence			0.838
No	269 (73.7%)	207 (74.7%)	
Yes	96 (26.3%)	70 (25.3%)	
Follow-up time	51.6 [23.0; 75.0]	15.0 [7.00; 22.0]	<0.001

excluded as they did not meet the inclusion criteria. Consequently, the analysis finally included 642 patients who were allocated to the training cohort (n=365) and the temporal validation cohort (n=277). **Table 1** summarizes the baseline characteristics of the enrolled STS patients. In the training cohort, the median age is 51 years (IQR, 38.0-63.0), the median follow-up duration is 51.6 months (IQR, 23.0-75.0), and the recurrence rate is 26.3%. In the validation cohort, the median age is 58 years (IQR, 45.0-69.0), the median follow-up duration is 15.0 months (IQR, 7.0-22.0), and the recurrence rate is 25.3%.

Predictor selection

The training cohort had 14 variables included for predictor selection. By applying 11 machine learning algorithms and applying the WM, 11 feature subsets were generated, as detailed in Table S1.

Model development, validation, and evaluation

A total of 121 candidate models were developed based on different algorithm-feature subset combinations, and their predictive performance was further evaluated in the validation cohort.

Our first model screening was conducted with the use of the C-index. **Figure 2** shows the top 50 candidates according to average C-index and the full ranking for all 121 models is presented in Figure S1. The COXPH+RSF model (CRM) exhibits the highest average C-index of 0.837 in the validation cohorts among all models which indicates superior performance. In the training cohort, C-index of CRM is 0.923 (95% CI 0.878-0.935), in the cross-validated training cohort, it is 0.867 (95% CI 0.850-0.875), and in the temporal validation cohort, it is 0.807 (95% CI 0.765-0.819). The nine predictors used in the CRM are: Age, chemotherapy, chemotherapy sensitive subtype, depth, grade, margin, N stage, pathology subtype, and size.

The time-dependent calibration curves for the 1, 3, and 5-year demonstrate that CRM achieves good calibration in the training cohort and the validation cohorts (**Figures 3** and S2). The model was well calibrated in the training cohort but it progressively overestimated recurrence-free survival over time in the temporal validation cohort, suggesting potential model overfitting and temporal distribution shift.

The time-dependent ROC curves show that the CRM maintains strong predictive accuracy at 1, 3, and 5 years in the training cohort and the validation cohorts. In the training cohort, CRM achieves an AUC of 0.941 (95% CI 0.902-0.972) at 1 year, 0.976 (95% CI 0.959-0.990) at 3 years, and 0.970 (95% CI 0.949-0.986) at 5 years (**Figure 4A**). In the validation cohort, CRM achieves an AUC of 0.863 (95% CI 0.792-0.929) at 1 year, 0.884 (95% CI 0.821-0.934) at 3 years, and 0.911 (95% CI 0.846-0.958) at 5 years (**Figure 4B**).

The DCA of the CRM demonstrates a constant net benefit in the training cohort and time validation cohorts over a range of threshold probabilities (Figure S3). The CRM is found to be superior to the 'treat none' and 'treat all' strategies in two cohorts which indicate its practical utility when making decisions.

Model interpretation and development of new stage system

The time-dependent feature importance curves show the varying importance of each predictor in CRM over time (**Figure 5**). It can be seen that margin is the most influential predictor of CRM throughout the follow-up period.

To make it easier to implement clinically, we have created a web-based risk calculator (https://psxuliu.shinyapps.io/COXRSFforSTS/).

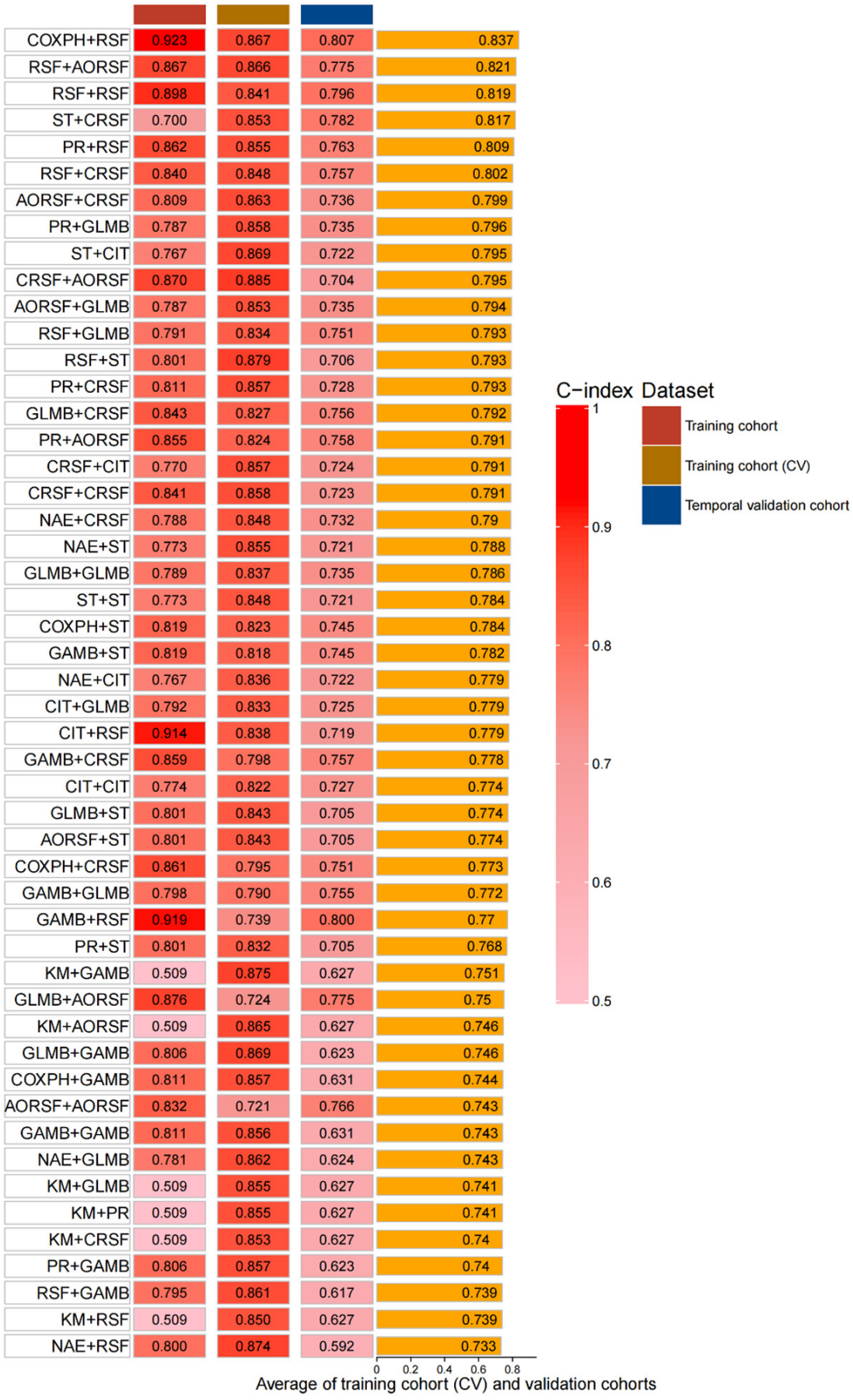

COXPH+RSF 0.923 0.867 0.807 0.837
RSF+AORSF 0.867 0.866 0.775 0.821
RSF+RSF 0.898 0.841 0.796 0.819
ST+CRSF 0.700 0.853 0.782 0.817
PR+RSF 0.862 0.855 0.763 0.809
RSF+CRSF 0.840 0.848 0.757 0.802
AORSF+CRSF 0.809 0.863 0.736 0.799
PR+GLMB 0.787 0.858 0.735 0.796
ST+CIT 0.767 0.869 0.722 0.795
CRSF+AORSF 0.870 0.885 0.704 0.795
AORSF+GLMB 0.787 0.853 0.735 0.794
RSF+GLMB 0.791 0.834 0.751 0.793
RSF+ST 0.801 0.879 0.706 0.793
PR+CRSF 0.811 0.857 0.728 0.793
GLMB+CRSF 0.843 0.827 0.756 0.792
PR+AORSF 0.855 0.824 0.758 0.791
CRSF+CIT 0.770 0.857 0.724 0.791
CRSF+CRSF 0.841 0.858 0.723 0.791
NAE+CRSF 0.788 0.848 0.732 0.79
NAE+ST 0.773 0.855 0.721 0.788
GLMB+GLMB 0.789 0.837 0.735 0.786
ST+ST 0.773 0.848 0.721 0.784
COXPH+ST 0.819 0.823 0.745 0.784
GAMB+ST 0.819 0.818 0.745 0.782
NAE+CIT 0.767 0.836 0.722 0.779
CIT+GLMB 0.792 0.833 0.725 0.779
CIT+RSF 0.914 0.838 0.719 0.779
GAMB+CRSF 0.859 0.798 0.757 0.778
CIT+CIT 0.774 0.822 0.727 0.774
GLMB+ST 0.801 0.843 0.705 0.774
AORSF+ST 0.801 0.843 0.705 0.774
COXPH+CRSF 0.861 0.795 0.751 0.773
GAMB+GLMB 0.798 0.790 0.755 0.772
GAMB+RSF 0.919 0.739 0.800 0.77
PR+ST 0.801 0.832 0.705 0.768
KM+GAMB 0.509 0.875 0.627 0.751
GLMB+AORSF 0.876 0.724 0.775 0.75
KM+AORSF 0.509 0.865 0.627 0.746
GLMB+GAMB 0.806 0.869 0.623 0.746
COXPH+GAMB 0.811 0.857 0.631 0.744
AORSF+AORSF 0.832 0.721 0.766 0.743
GAMB+GAMB 0.811 0.856 0.631 0.743
NAE+GLMB 0.781 0.862 0.624 0.743
KM+GLMB 0.509 0.855 0.627 0.741
KM+PR 0.509 0.855 0.627 0.741
KM+CRSF 0.509 0.853 0.627 0.74
PR+GAMB 0.806 0.857 0.623 0.74
RSF+GAMB 0.795 0.861 0.617 0.739
KM+RSF 0.509 0.850 0.627 0.739
NAE+RSF 0.800 0.874 0.592 0.733
0 0.2 0.4 0.6 0.8
Average of training cohort (CV) and validation cohorts
C−index
1
0.9
0.8
0.7
0.6
0.5
Dataset
Training cohort
Training cohort (CV)
Temporal validation cohort

Figure 2. Concordance index of top 50 machine learning models. The C-index for the top 50 out of 121 machine learning models was calculated for the training cohort and three validation cohorts. Ranking of the models was based on the average C-index of three validation cohorts. GAMB, Boosted Generalized Additive Model; GLMB, Boosted Generalized Linear Model; ST, Survival Tree; CIT, Conditional Inference Tree; RSF, Random Survival Forest; CRF, Conditional Random Forest; AORSF, Accelerated Oblique Random Survival Forest; PR, Penalized Regression; COXPH, Cox Proportional Hazards Model; KM, Kaplan-Meier Estimator; NAE, Normalized Absolute Error.

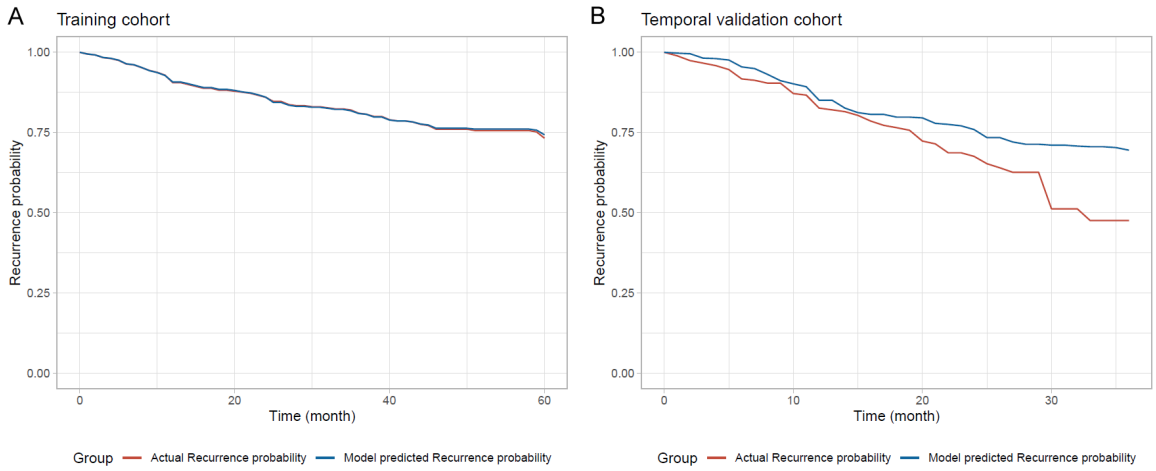


Figure 3. Evaluating the calibration of CRM by time-dependent calibration curves. A. Time-dependent calibration curve of training cohort; B. Time-dependent calibration curve of validation cohort.

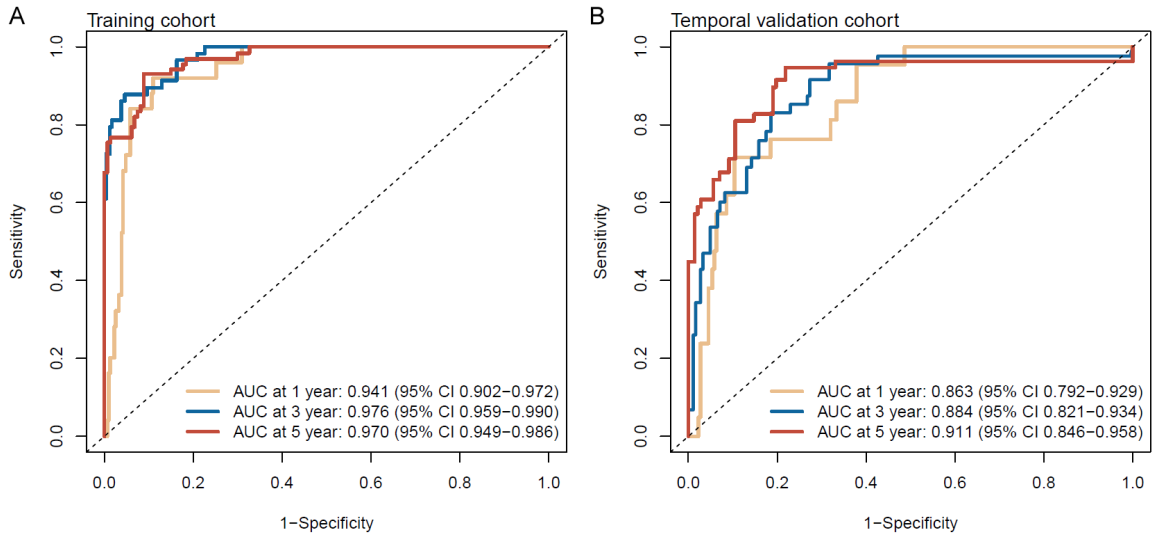


Figure 4. Evaluating the predictive accuracy of CRM by time-dependent ROC curves. A. Time-dependent ROC curves training cohort; B. Time-dependent ROC curves temporal validation cohort. CRM, Cox Proportional Hazards Model + Random Survival Forest.

Also, we developed a staging system based on CRM and using the ST algorithm to divide STS patients into two risk groups by risk scores. The cumulative risk curves indicate that there are statistically significant differences in recurrence rates between the high-risk and low-risk groups in all four cohorts (Figure S4A, S4B).

The significance of the CRM stage system in guiding postoperative radiotherapy

We used the CRM staging system to divide patients in both cohorts into high-risk and low-risk groups. Cumulative risk curves were then generated to compare the outcomes of those who received adjuvant radiotherapy against

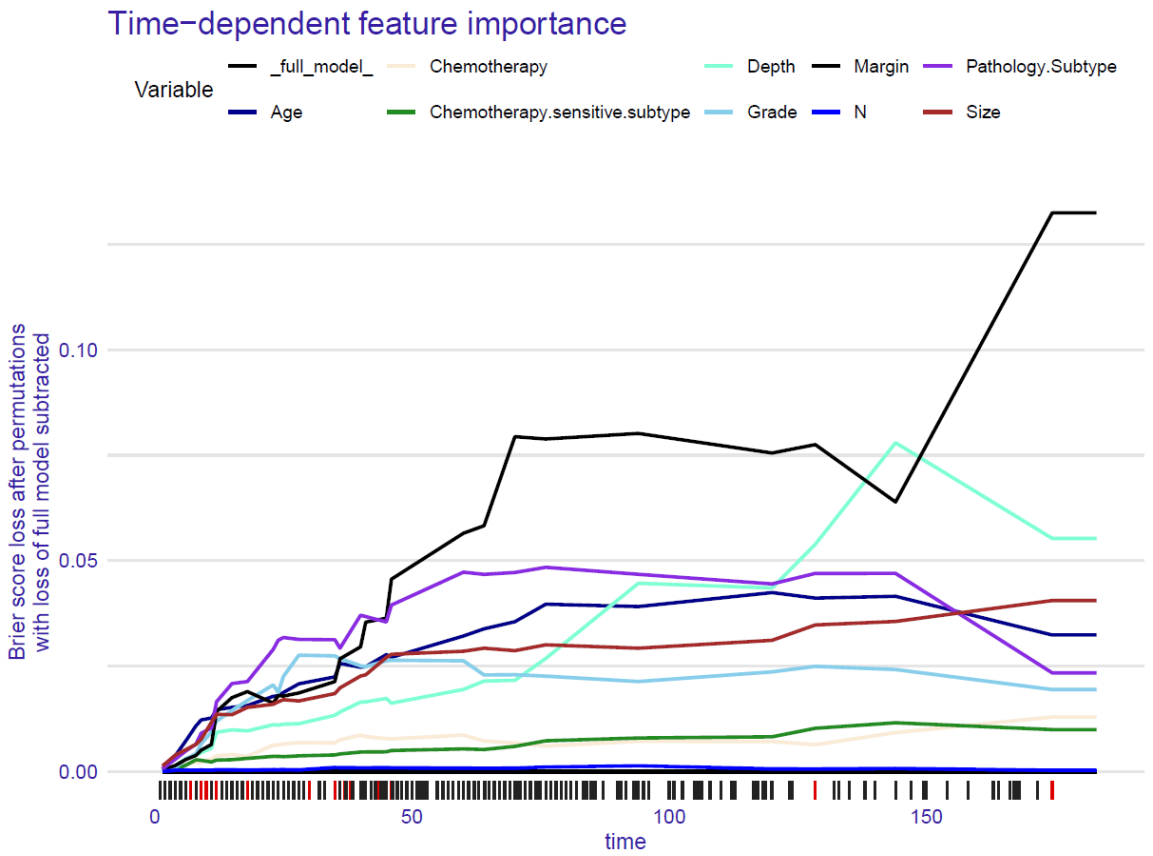


Figure 5. Interpretation the CRM by time-dependent feature importance curves and the performance of the CRM staging system in the training cohort and validation cohort. The time-dependent feature importance curves of CRM.

those who did not. The results show that, in the high-risk groups of both cohorts, the differences in recurrence rates between the radiotherapy and non-radiotherapy subgroups are statistically significant (**Figure 6A**, **6C**). However, in the low-risk groups, although a trend toward different recurrence rates is observed between the two subgroups, the differences do not reach statistical significance (**Figure 6B**, **6D**). These findings suggest that the CRM staging system can effectively evaluate the benefits of postoperative radiotherapy and provide valuable guidance for adjuvant treatment decision-making.

Discussion

This research developed and validated a machine learning-based model to predict postoperative recurrence risk in STS patients. The resultant CRM was found to be more accurate in predicting the outcome in both the training cohort and the temporal validation cohort. It achieved higher C-index and AUC values compared to other models, suggesting better accuracy and stability in recurrence prediction. The time-dependent calibration curves and DCA reveal the superior calibration and clinical net benefit. Such results indicate that the CRM has the potential to identify postoperative recurrence in STS patients. It may also help clinicians evaluate the severity of the disease, facilitate patient follow-up, and provide directions on multidisciplinary treatment strategies. Combining the CRM with the ST staging system provides additional guidance for postoperative adjuvant radiotherapy in STS patients.

The recent years have seen an increased interest in the creation of prediction models in clinical research. It emphasizes the importance of adhering to standardized procedures for model construction and validation. The present study was strictly aligned with the existing standards and carefully assessed to guarantee methodological integrity. As indicated by earlier studies, the rapid development of predictive modeling has been accompanied by increasing demands for transparency and accuracy. The pub-

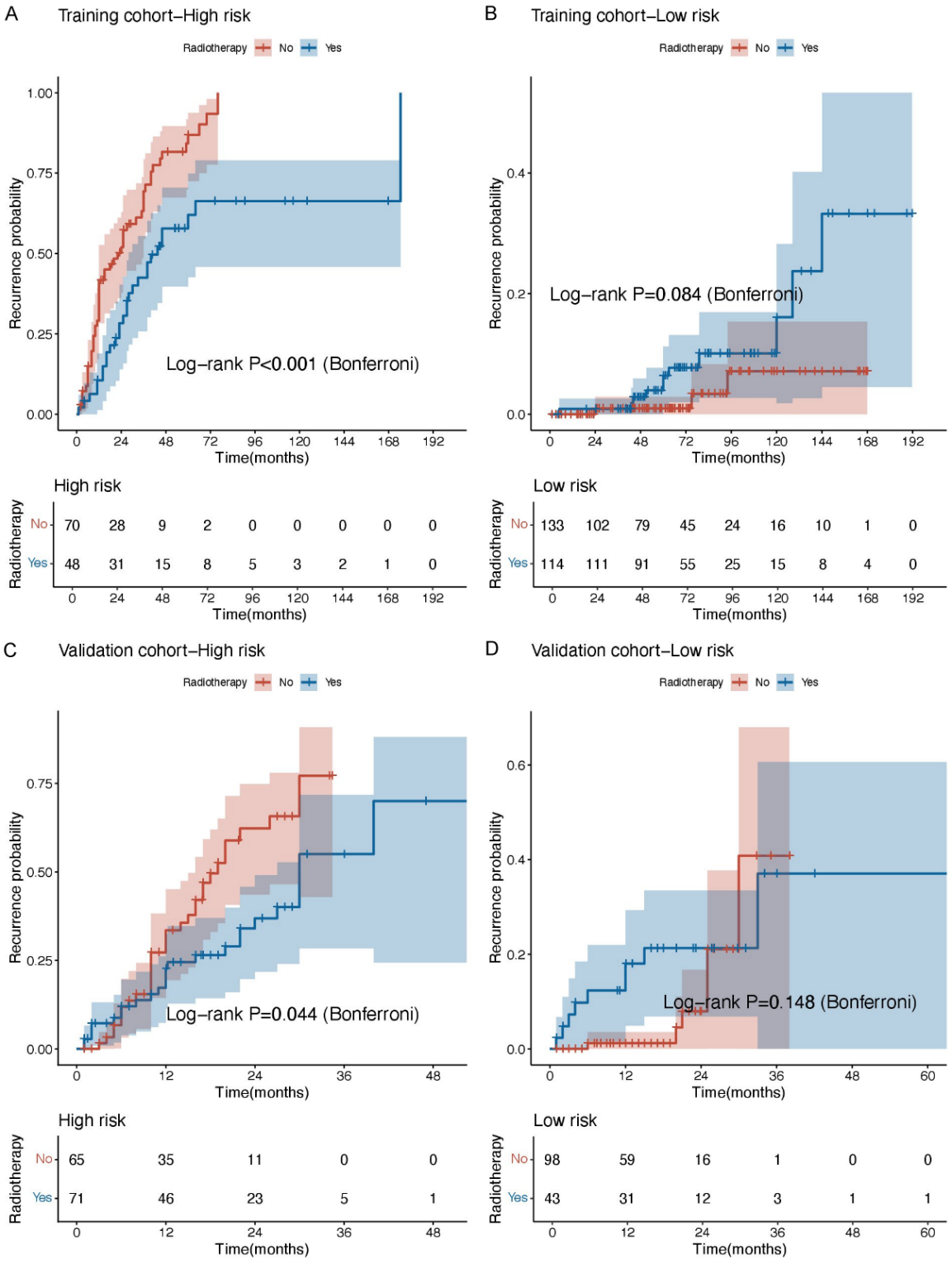


Figure 6. The significance of CRM staging in guiding postoperative radiotherapy. A. Cumulative risk curves of the high-risk group based on the CRM staging system in the training cohort; B. Cumulative risk curves of the low-risk group based on the CRM staging system in the training cohort; C. Cumulative risk curves of the high-risk group based on the CRM staging system in the validation cohort; D. Cumulative risk curves of the low-risk group based on the CRM staging system in the validation cohort. CRM, Cox proportional hazards model + random survival forest.

lication of the TRIPOD statement marked an important step forward in this area [19, 23]. The assessment tool proposed by Wolff et al. since then has provided a systematic structure to assess bias and applicability [17]. The checklist introduced by Markowetz has further emphasized the feasibility of the prediction models in clinical practice [18]. In this study, we followed these established tools and recommendations and adopted a standardized research workflow.

Postoperative adjuvant radiotherapy plays an important role in the multidisciplinary management of soft tissue sarcoma and contributes to better local control [24]. Nevertheless, postoperative adjuvant treatment strategies have not yet been widely standardized. Its therapeutic efficacy and safety remain under investigation. The CRM developed in our research may provide a useful basis for the development of postoperative adjuvant treatment strategies in patients with STS. In resource-constrained healthcare settings, the CRM can improve resource allocation by identifying high-risk patients for intensified management while helping low-risk patients avoid unnecessary interventions.

Older age portends worse outcomes. This may be related to frailty, comorbidity and immunosenescence, which may reduce tolerance to multimodality therapy and impair antitumor immunity. This can lead to earlier treatment de-escalation and micrometastatic escape [25, 26]. Adjuvant chemotherapy can improve survival primarily by eradicating occult micrometastases in high-risk resected disease, although its benefit varies across regimens and histotypes [27]. Depth of tumor measures anatomical boundaries and pathways of extension. Deeply located lesions track along fascial compartments and critical structures, raising the risk of positive surgical margins and vascular invasion and subsequently lead to local recurrence and distant metastasis [28, 29]. Histological grade, particularly a high FNCLCC grade, is indicative of more aggressive tumor biology and is associated with poorer survival outcomes and is more prone to distant recurrence [25]. Surgical margin status critically impacts local control. Negative margins greater than 1 mm are associated with a substantially lower risk of local recurrence and negative margins above 5 mm offer optimal 5-year local-recurrence-free survival [30]. Regional lymph node involvement suggests a biologically aggressive phenotype that can spread both in the form of lymphovascular and hematogenous spread. Its presence is associated with a dramatic decline in both 6-year overall survival and disease-free survival [31]. Histological subtype is an independent determinant of clinical outcomes. Well-differentiated liposarcoma in liposarcoma is associated with good survival (5-year rates exceeding 82%), whereas dedifferentiated and pleomorphic subtypes are linked to poor prognoses [29, 32]. Tumor size one of the most significant determinants of prognosis in soft tissue sarcoma. Tumors larger than 5 cm are more likely to develop hypoxia, necrosis and to exhibit early micrometastatic spread, which is associated with worse survival [25, 29, 33].

This study employs a single-center training cohort and a temporal validation cohort. This design helps maintain data consistency and it is easier to handle when it comes to privacy and costs. The uniform clinical environment and standardized data collection procedures make data preprocessing and feature engineering easier and less likely to be affected by heterogeneity. The analysis and model development are also made easier by using data from only a single institution since ethical and privacy monitoring can be done more easily. Temporal validation allows assessment of the model's stability across patients from different time periods and may also help identify possible temporal drift and thus improve predictive reliability [34].

Despite these advantages, the study inherently presents several limitations. It casts doubt on the model generalizability since the model may overfit institution-specific patient characteristics, and thus the predictions could be biased. The model development and validation in our research is primarily based on retrospective data [35], additional evidence in prospective studies should be done. Given the biological heterogeneity of STS subtypes, further investigation is needed to determine whether the model can be applied to different subgroups. To enhance reliability, generalizability, and clinical translational value of the model, future validation involving multicenter, large-scale and

more representative cohorts is inevitable to enhance the model further.

There was a small calibration drift in the validation cohort with later time points. It is probably because the temporal validation cohort has a shorter follow-up period, meaning that less patients are at risk and there is more censoring late on. Due to the small quantity of long-term occurrences, the stability of observed probabilities might be decreased and apparent inconsistencies in calibration can occur. In particular, the model demonstrated a good correspondence of the predicted and observed results throughout the clinically meaningful follow-up period, which confirms the validity of the model to predict short to mid-term results, although it should be used with caution when extrapolated into the long term.

Conclusion

In summary, we developed a machine learning model (CRM) to predict recurrence in STS patients with great precision after radical surgery. The CRM demonstrates high predictive performance, good calibration, and consistent clinical net benefit in an independent validation cohort. Due to its high accuracy and reliability, the CRM may be used as an effective tool to predict postoperative recurrence and guide adjuvant treatment strategies in STS patients.

Acknowledgements

We express our gratitude to all the developers of the R programming package for generously sharing their code. This study was supported by the National Natural Science Foundation of China (No. 82272964) and National High Level Hospital Clinical Research Funding (No. 80102022525).

Disclosure of conflict of interest

None.

Address correspondence to: Shengji Yu, Department of Orthopedics, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 17, Panjiayuan South Lane, Chaoyang District, Beijing 100021, China. Tel: +86-010-87788722; ORCID: 0000-0002-8994-0562; E-mail: zlyygk@163.com; shengjiyu@126.com

References

[1] Liu X, Yuan J, Wang X and Yu S. Development and validation of a machine learning-based model to predict postoperative overall survival in patients with soft tissue sarcoma: a retrospective cohort study. Am J Cancer Res 2024; 14: 4731-4746.

[2] Liu H, Zhang H, Zhang C, Liao Z, Li T, Yang T, Zhang G and Yang J. Pan-soft tissue sarcoma analysis of the incidence, survival, and metastasis: a population-based study focusing on distant metastasis and lymph node metastasis. Front Oncol 2022; 12: 890040.

[3] Acem I and van de Sande MAJ. Prediction tools for the personalized management of soft-tissue sarcomas of the extremity. Bone Joint J 2022; 104-B: 1011-1016.

[4] Callegaro D, Sarre Lazcano C and Cardona K. ASO practice guidelines series: soft tissue sarcoma of the extremities and superficial trunk. Ann Surg Oncol 2025; [Epub ahead of print].

[5] Von Mehren M, Kane JM, Agulnik M, Bui MM, Carr-Ascher J, Choy E, Connelly M, Dry S, Ganjoo KN, Gonzalez RJ, Holder A, Homsi J, Keedy V, Kelly CM, Kim E, Liebner D, McCarter M, McGarry SV, Mesko NW, Meyer C, Pappo AS, Parkes AM, Petersen IA, Pollack SM, Poppe M, Riedel RF, Schuetze S, Shabason J, Sicklick JK, Spraker MB, Zimel M, Hang LE, Sundar H and Bergman MA. Soft tissue sarcoma, version 2.2022, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2022; 20: 815-833.

[6] Hayes AJ, Nixon IF, Strauss DC, Seddon BM, Desai A, Benson C, Judson IR and Dangoor A. UK guidelines for the management of soft tissue sarcomas. Br J Cancer 2025; 132: 11-31.

[7] Delisle M, Gyorki D, Bonvalot S and Nessim C. Landmark series: a review of landmark studies in the treatment of primary localized retroperitoneal sarcoma. Ann Surg Oncol 2022; 29: 7297-7311.

[8] Yang X, Zhang L, Yang X, Yu W and Fu J. Oncologic outcomes of pre- versus post-operative radiation in resectable soft tissue sarcoma: a systematic review and meta-analysis. Radiat Oncol 2020; 15: 158.

[9] Díaz Casas SE, Villacrés JM, Lehmann Mosquera C, García Mora M, Mariño Lozano I, Ángel Aristizábal J, Suarez Rodríguez R, Duarte Torres CA and Sánchez Pedraza R. Prognostic factors associated with tumor recurrence and overall survival in soft tissue sarcomas of the extremities in a Colombian reference cancer center. Curr Oncol 2024; 31: 1725-1738.

[10] Yuan J, Liu X, Zhao Z and Yu S. Enhancing the accuracy of survival prediction for synovial sarcoma: a comparative study of machine learn-

ing models. Asian J Surg 2024; S1015-9584(24)02773-8: [Epub ahead of print].
[11] Liu X, Xiao Q, Gu Z, Wu X, Yuan C, Tang X, Meng F, Wang D, Lang R, Guo K, Tian X, Zhang Y, Zhao E, Wu Z, Xu J, Xing Y, Cao F, Wang C and Zhang J. Development and external validation of a machine learning-based model to predict postoperative recurrence in patients with duodenal adenocarcinoma: a multicenter, retrospective cohort study. BMC Med 2025; 23: 98.
[12] Chen Q, Chen J, Deng Y, Bi X, Zhao J, Zhou J, Huang Z, Cai J, Xing B, Li Y, Li K and Zhao H. Personalized prediction of postoperative complication and survival among colorectal liver metastases patients receiving simultaneous resection using machine learning approaches: a multi-center study. Cancer Lett 2024; 593: 216967.
[13] Margue G, Ferrer L, Etchepare G, Bigot P, Bensalah K, Mejean A, Roupret M, Doumerc N, Ingels A, Boissier R, Pignot G, Parier B, Paparel P, Waeckel T, Colin T and Bernhard JC. UroPredict: machine learning model on real-world data for prediction of kidney cancer recurrence (UroCCR-120). NPJ Precis Oncol 2024; 8: 45.
[14] Zhang Y, Yang Z, Chen R, Zhu Y, Liu L, Dong J, Zhang Z, Sun X, Ying J, Lin D, Yang L and Zhou M. Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer. NPJ Digit Med 2024; 7: 15.
[15] Fu M, Lin Y, Yang J, Cheng J, Lin L, Wang G, Long C, Xu S, Lu J, Li G, Yan J, Chen G, Zhuo S and Chen D. Multitask machine learning-based tumor-associated collagen signatures predict peritoneal recurrence and disease-free survival in gastric cancer. Gastric Cancer 2024; 27: 1242-1257.
[16] Chen S, Li N, Tang Y, Chen B, Fang H, Qi S, Lu N, Yang Y, Song Y, Liu Y, Wang S, Li YX and Jin J. Radiomics analysis of fat-saturated T2-weighted MRI sequences for the prediction of prognosis in soft tissue sarcoma of the extremities and trunk treated with neoadjuvant radiotherapy. Front Oncol 2021; 11: 710649.
[17] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J and Mallett S; PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019; 170: 51-58.
[18] Markowetz F. All models are wrong and yours are useless: making clinical prediction models impactful for patients. NPJ Precis Oncol 2024; 8: 54.
[19] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF and Collins GS. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015; 162: W1-W73.
[20] Sonabend R, Király FJ, Bender A, Bischl B and Lang M. mlr3proba: an R package for machine learning in survival analysis. Bioinformatics 2021; 37: 2789-2791.
[21] Gonzalez-Dias P, Lee EK, Sorgi S, de Lima DS, Urbanski AH, Silveira EL and Nakaya HI. Methods for predicting vaccine immunogenicity and reactogenicity. Hum Vaccin Immunother 2020; 16: 269-276.
[22] Xiao Z and Lam HM. ShinySyn: a Shiny/R application for the interactive visualization and integration of macro- and micro-synteny data. Bioinformatics 2022; 38: 4406-4408.
[23] Fihn SD, Berlin JA, Haneuse SJPA and Rivara FP. Prediction models and clinical outcomes - a call for papers. JAMA Netw Open 2024; 7: e249640.
[24] Sunyach MP, Lusque A, Le Péchoux C, Levy A, Sargos P, Helfre S, Thariat J, Moureau Zabotto L, Lerouge D, Llacer C, Mervoyer A, Vogin G, Chevreau C, Ducimetière F, Blay JY, Delannes M and Ducassou A. Postoperative radiotherapy in patients with R0 resection of soft tissue sarcoma: results from the European sarcoma CONTICABASE analysis. Br J Radiol 2025; 98: 1409-1418.
[25] Zastrow RK, El Sayed M, LiBrizzi CL, Jacobs AJ and Levin AS. Progressive improvement in 5-year survival rates for extremity soft tissue sarcomas from 1999 to 2019. Sarcoma 2024; 2024: 8880609.
[26] Zhang D, Hu J, Liu Z, Wu H, Cheng H and Li C. Prognostic nomogram in patients with epithelioid sarcoma: a SEER-based study. Cancer Med 2023; 12: 3079-3088.
[27] Goh MH, Gonzalez MR, Heiling HM, Mazzola E, Connolly JJ, Choy E, Cote GM, Spentzos D and Lozano-Calderon SA. Adjuvant chemotherapy in localized, resectable extremity and truncal soft tissue sarcoma and survival outcomes - a systematic review and meta-analysis of randomized controlled trials. Cancer 2025; 131: e35792.
[28] Lin JS, Coleman L, Voskuil RT, Malik A, Mayerson JL and Scharschmidt TJ. Local recurrence rates of superficial versus deep soft tissue sarcoma. Arch Orthop Trauma Surg 2024; 144: 2967-2973.
[29] Li RH, Zhou Q, Li AB, Zhang HZ and Lin ZQ. A nomogram to predict metastasis of soft tissue sarcoma of the extremities. Medicine (Baltimore) 2020; 99: e20165.
[30] Steffens JM, Budny T, Gosheger G, De Vaal M, Rachbauer AM, Laufer A, Engel NM and Deventer N. The impact of resection margins in primary resection of high-grade soft tissue

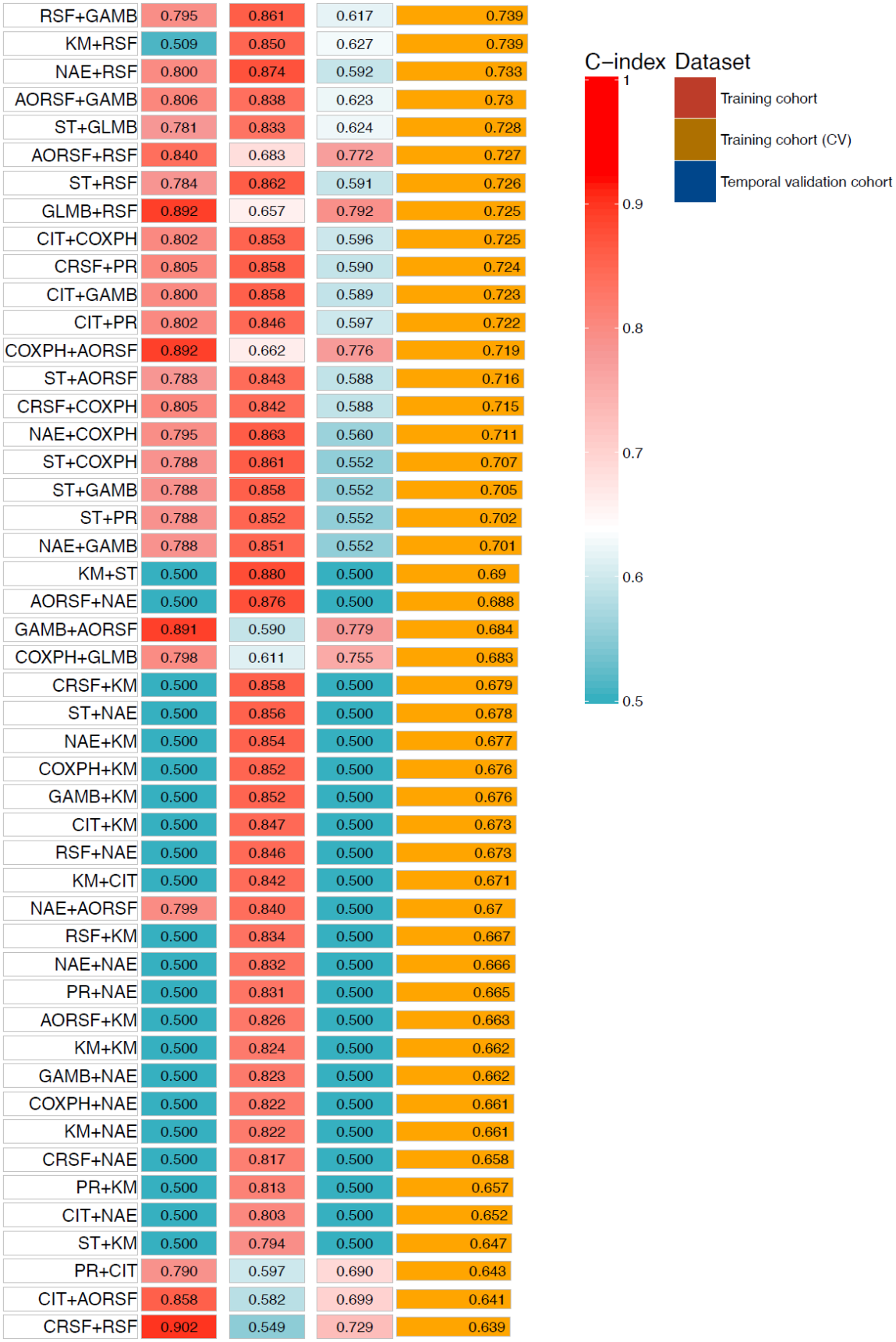
sarcomas: how far is far enough? Biomedicines 2025; 13: 1011.
[31] Shalaby M, Allam RM, Elkordy MA and Taher M. Prognostic significance of lymph node metastasis of soft tissue sarcoma of the extremities. National cancer institute experience. Int J Clin Oncol 2025; 30: 407-416.
[32] Zhao J, Du W, Tao X, Li A, Li Y and Zhang S. Survival and prognostic factors among different types of liposarcomas based on SEER database. Sci Rep 2025; 15: 1790.
[33] Isaac C, Kavanagh J, Griffin AM, Dickie CI, Mohankumar R, Chung PW, Catton CN, Shultz D, Ferguson PC and Wunder JS. Radiological progression of extremity soft tissue sarcoma following pre-operative radiotherapy predicts for poor survival. Br J Radiol 2022; 95: 20210936.
[34] Meinert CL and Tonascia S. Single-center versus multicenter trials. In: Meinert CL, editors. Clinical trials: design, conduct and analysis. Oxford University Press; 1986.
[35] Samaga D, Hornung R, Braselmann H, Hess J, Zitzelsberger H, Belka C, Boulesteix AL and Unger K. Single-center versus multi-center data sets for molecular prognostic modeling: a simulation study. Radiat Oncol 2020; 15: 109.

Table S1. Feature subsets after WM feature selection

Models	Feature subset
RSF	Margin; Pathology Subtype
GAMB	Age; Chemotherapy; Chemotherapy sensitive subtype; Depth; Grade; Margin; Pathology Subtype; Site
GLMB	Chemotherapy; Chemotherapy sensitive subtype; Depth; Grade; Margin; N stage; Pathology Subtype; Radiotherapy; Sex; Site
CIT	Age; Margin; Pathology Subtype; Size
CRF	Depth; Margin; Pathology Subtype; Sex; Site; Size
COXPH	Age; Chemotherapy; Chemotherapy sensitive subtype; Depth; Grade; Margin; N stage; Pathology Subtype; Size
KM	N stage
NAE	Chemotherapy; Margin; Pathology Subtype
AORSF	Depth; Grade; Margin; Pathology Subtype; Radiotherapy
PR	Chemotherapy sensitive subtype; Grade; Margin; Pathology Subtype; Site; Size
ST	Chemotherapy; Depth; Grade; Margin; Pathology Subtype; Sex

WM, Wrapper method; RSF, Random Survival Forest; GAMB, Generalized Additive Models via Gradient Boosting; GLMB, Generalized Linear Models via Gradient Boosting; CIT, Conditional Inference Tree; CRF, Conditional Random Forest; COXPH, Cox Proportional Hazards Model; KM, Kaplan-Meier Estimator; NAE, Normalized Absolute Error; AORSF, Accelerated Oblique Random Survival Forest; PR, Penalized Regression; ST, Survival Tree.

COXPH+RSF	0.923	0.867	0.807	0.837
RSF+AORSF	0.867	0.866	0.775	0.821
RSF+RSF	0.898	0.841	0.796	0.819
ST+CRSF	0.700	0.853	0.782	0.817
PR+RSF	0.862	0.855	0.763	0.809
RSF+CRSF	0.840	0.848	0.757	0.802
AORSF+CRSF	0.809	0.863	0.736	0.799
PR+GLMB	0.787	0.858	0.735	0.796
ST+CIT	0.767	0.869	0.722	0.795
CRSF+AORSF	0.870	0.885	0.704	0.795
AORSF+GLMB	0.787	0.853	0.735	0.794
RSF+GLMB	0.791	0.834	0.751	0.793
RSF+ST	0.801	0.879	0.706	0.793
PR+CRSF	0.811	0.857	0.728	0.793
GLMB+CRSF	0.843	0.827	0.756	0.792
PR+AORSF	0.855	0.824	0.758	0.791
CRSF+CIT	0.770	0.857	0.724	0.791
CRSF+CRSF	0.841	0.858	0.723	0.791
NAE+CRSF	0.788	0.848	0.732	0.79
NAE+ST	0.773	0.855	0.721	0.788
GLMB+GLMB	0.789	0.837	0.735	0.786
ST+ST	0.773	0.848	0.721	0.784
COXPH+ST	0.819	0.823	0.745	0.784
GAMB+ST	0.819	0.818	0.745	0.782
NAE+CIT	0.767	0.836	0.722	0.779
CIT+GLMB	0.792	0.833	0.725	0.779
CIT+RSF	0.914	0.838	0.719	0.779
GAMB+CRSF	0.859	0.798	0.757	0.778
CIT+CIT	0.774	0.822	0.727	0.774
GLMB+ST	0.801	0.843	0.705	0.774
AORSF+ST	0.801	0.843	0.705	0.774
COXPH+CRSF	0.861	0.795	0.751	0.773
GAMB+GLMB	0.798	0.790	0.755	0.772
GAMB+RSF	0.919	0.739	0.800	0.77
PR+ST	0.801	0.832	0.705	0.768
KM+GAMB	0.509	0.875	0.627	0.751
GLMB+AORSF	0.876	0.724	0.775	0.75
KM+AORSF	0.509	0.865	0.627	0.746
GLMB+GAMB	0.806	0.869	0.623	0.746
COXPH+GAMB	0.811	0.857	0.631	0.744
AORSF+AORSF	0.832	0.721	0.766	0.743
GAMB+GAMB	0.811	0.856	0.631	0.743
NAE+GLMB	0.781	0.862	0.624	0.743
KM+GLMB	0.509	0.855	0.627	0.741
KM+PR	0.509	0.855	0.627	0.741
KM+CRSF	0.509	0.853	0.627	0.74
PR+GAMB	0.806	0.857	0.623	0.74


RSF+GAMB 0.795 0.861 0.617 0.739
KM+RSF 0.509 0.850 0.627 0.739
NAE+RSF 0.800 0.874 0.592 0.733
AORSF+GAMB 0.806 0.838 0.623 0.73
ST+GLMB 0.781 0.833 0.624 0.728
AORSF+RSF 0.840 0.683 0.772 0.727
ST+RSF 0.784 0.862 0.591 0.726
GLMB+RSF 0.892 0.657 0.792 0.725
CIT+COXPH 0.802 0.853 0.596 0.725
CRSF+PR 0.805 0.858 0.590 0.724
CIT+GAMB 0.800 0.858 0.589 0.723
CIT+PR 0.802 0.846 0.597 0.722
COXPH+AORSF 0.892 0.662 0.776 0.719
ST+AORSF 0.783 0.843 0.588 0.716
CRSF+COXPH 0.805 0.842 0.588 0.715
NAE+COXPH 0.795 0.863 0.560 0.711
ST+COXPH 0.788 0.861 0.552 0.707
ST+GAMB 0.788 0.858 0.552 0.705
ST+PR 0.788 0.852 0.552 0.702
NAE+GAMB 0.788 0.851 0.552 0.701
KM+ST 0.500 0.880 0.500 0.69
AORSF+NAE 0.500 0.876 0.500 0.688
GAMB+AORSF 0.891 0.590 0.779 0.684
COXPH+GLMB 0.798 0.611 0.755 0.683
CRSF+KM 0.500 0.858 0.500 0.679
ST+NAE 0.500 0.856 0.500 0.678
NAE+KM 0.500 0.854 0.500 0.677
COXPH+KM 0.500 0.852 0.500 0.676
GAMB+KM 0.500 0.852 0.500 0.676
CIT+KM 0.500 0.847 0.500 0.673
RSF+NAE 0.500 0.846 0.500 0.673
KM+CIT 0.500 0.842 0.500 0.671
NAE+AORSF 0.799 0.840 0.500 0.67
RSF+KM 0.500 0.834 0.500 0.667
NAE+NAE 0.500 0.832 0.500 0.666
PR+NAE 0.500 0.831 0.500 0.665
AORSF+KM 0.500 0.826 0.500 0.663
KM+KM 0.500 0.824 0.500 0.662
GAMB+NAE 0.500 0.823 0.500 0.662
COXPH+NAE 0.500 0.822 0.500 0.661
KM+NAE 0.500 0.822 0.500 0.661
CRSF+NAE 0.500 0.817 0.500 0.658
PR+KM 0.500 0.813 0.500 0.657
CIT+NAE 0.500 0.803 0.500 0.652
ST+KM 0.500 0.794 0.500 0.647
PR+CIT 0.790 0.597 0.690 0.643
CIT+AORSF 0.858 0.582 0.699 0.641
CRSF+RSF 0.902 0.549 0.729 0.639
C–index
1
0.9
0.8
0.7
0.6
0.5
Dataset
Training cohort
Training cohort (CV)
Temporal validation cohort

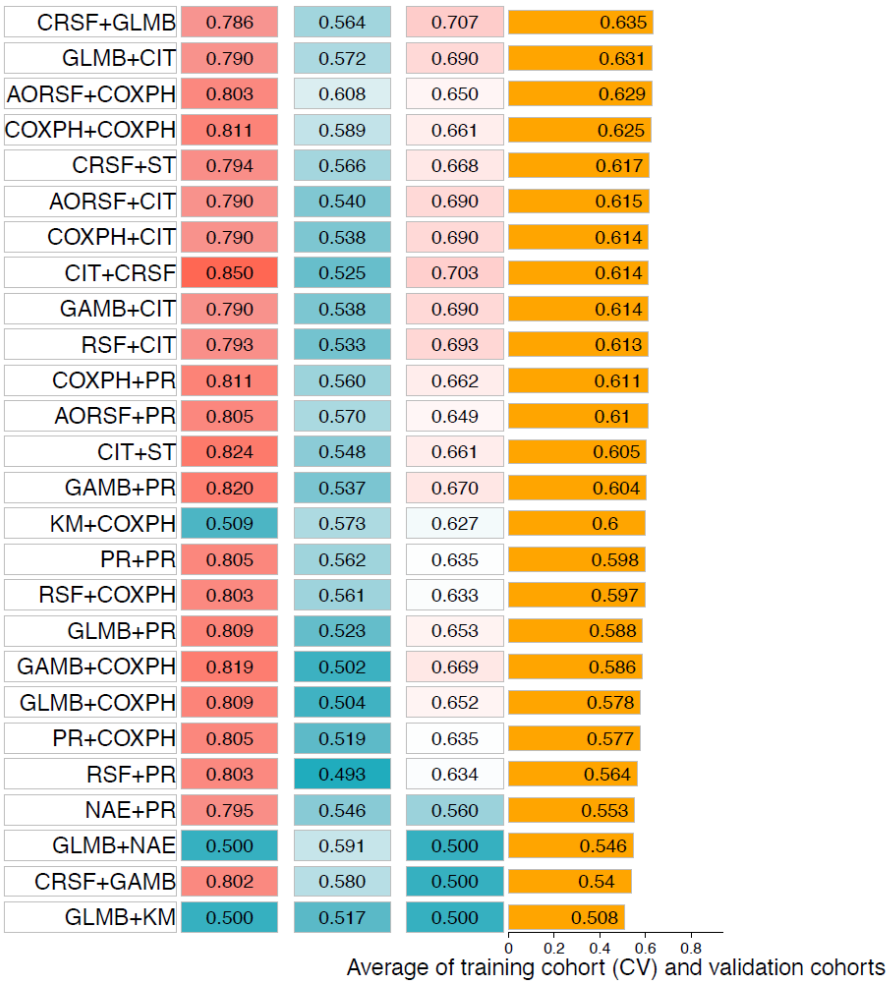


Figure S1. Concordance index of 121 machine learning models. The C-index for the 121 machine learning models was calculated for the training cohort and the validation cohort. Ranking of the models was based on the average C-index two cohorts.

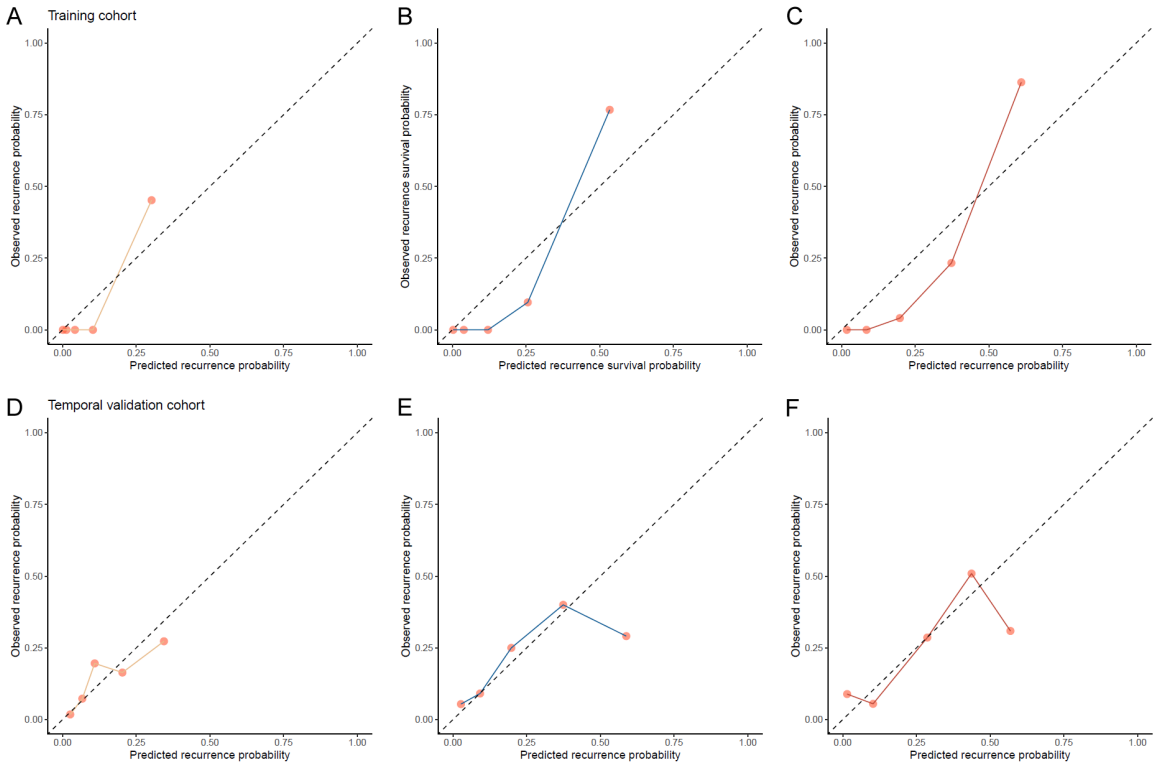


Figure S2. Evaluating the calibration of CRM by calibration curves. A. Calibration curves in 1-year in training cohort; B. Calibration curves in 3-year in training cohort; C. Calibration curves in 5-year in training cohort; D. Calibration curves in 1-year in validation cohort; E. Calibration curves in 3-year in validation cohort; F. Calibration curves in 5-year in validation cohort.

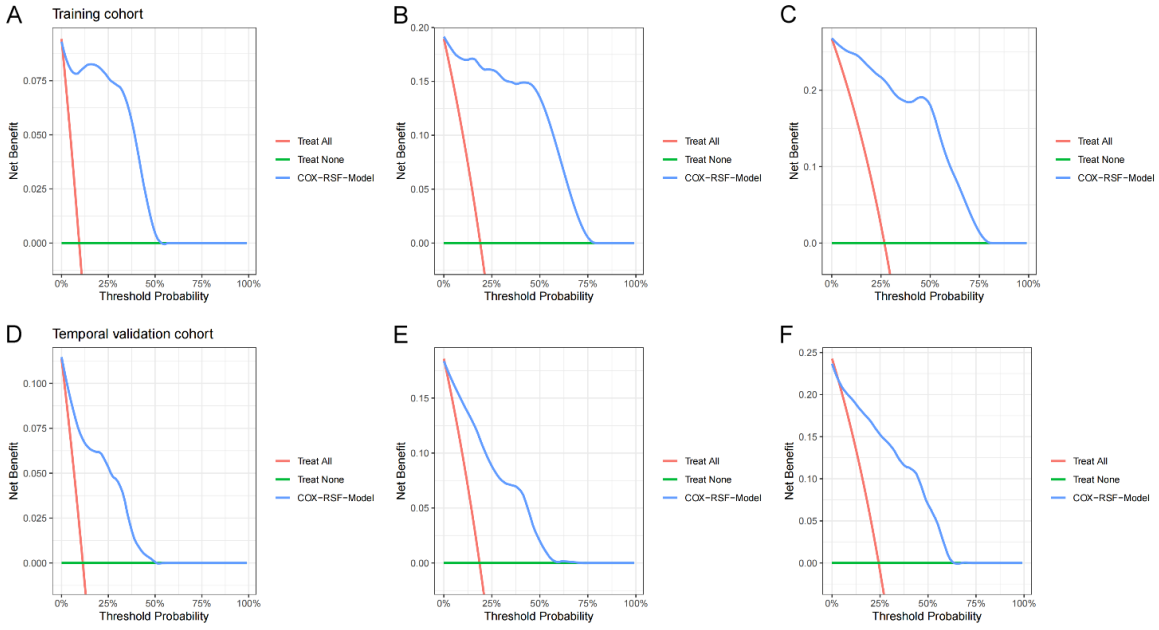


Figure S3. Evaluating the net benefit of CRM by DCA. A. 1-year DCA for training cohort; B. 3-year DCA for training cohort; C. 5-year DCA for training cohort; D. 1-year DCA for temporal validation cohort; E. 3-year DCA for temporal validation cohort; F. 5-year DCA for temporal validation cohort. DCA, Decision curve analysis.

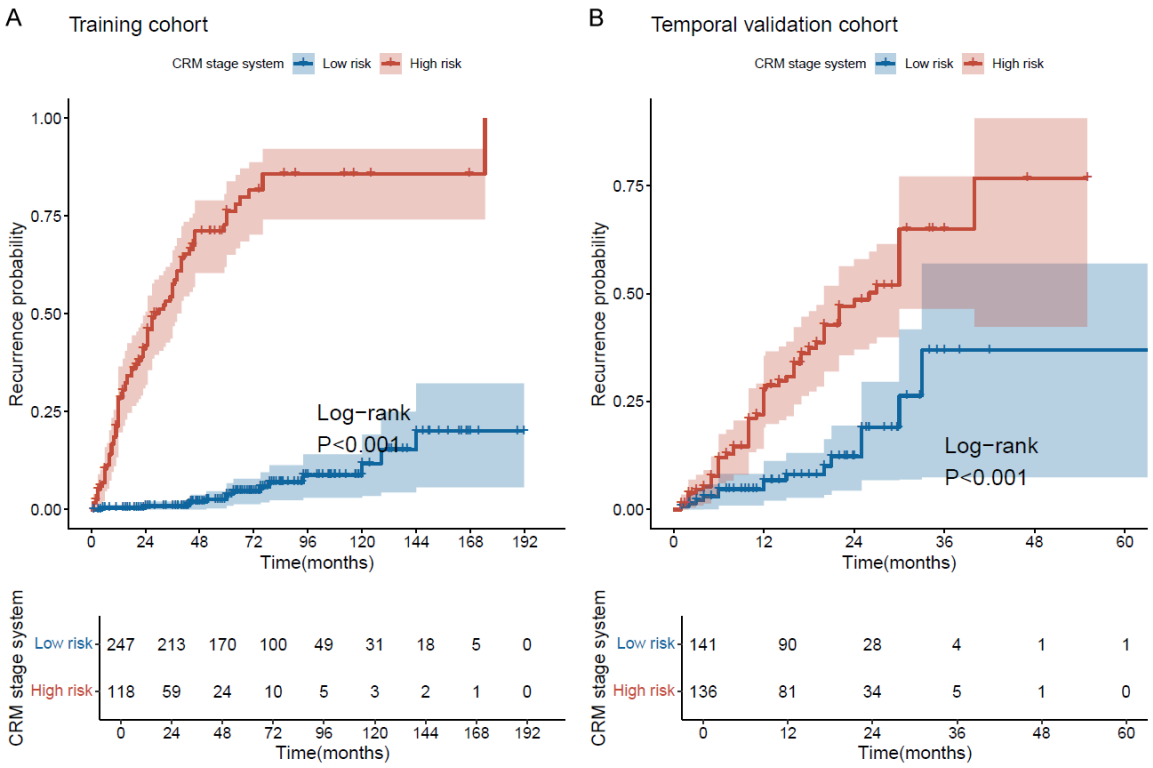


Figure S4. Performance of the CRM staging system in the training cohort and temporal validation cohort. A. Training cohort; B. Temporal validation cohort.