Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Construction and Cross-Cohort Validation of Survival Prediction Models for Prostate Cancer: An Interpretable Machine Learning Study Based on SEER and Local Chinese Clinical Data

Saimaitikari Abudoubari
The First People's Hospital of Kashi Prefecture
Aikebaierjiang Tuluhong
The First People's Hospital of Kashi Prefecture
Palidanmu Wumaier
The First People's Hospital of Kashi Prefecture
Jesur Batur
The First People's Hospital of Kashi Prefecture
Gulinigaer Wusiman
Bayingol Mongolian Autonomous Prefecture People's Hospital
Abudoula Tuerhong
The First People's Hospital of Kashi Prefecture
Maimaitiyiming Yasheng
The First People's Hospital of Kashi Prefecture
Ya Qiu
1254695790@qq.com

The First People's Hospital of Kashi Prefecture
Abudouresuli Tuersun
The First People's Hospital of Kashi Prefecture

Research Article

Keywords: Prostate cancer, Interpretable machine learning, Overall survival, SurvSHAP

Posted Date: August 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10588631/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objective

This investigation sought to develop interpretable machine learning(ML) frameworks for forecasting overall survival (OS) in prostate cancer patients, thereby enhancing prognostic transparency to facilitate individualized clinical decision-making.

Methods

Patient data were sourced from the SEER database(randomly partitioned into a 7:3 ratio for training and internal validation) and supplemented by an external cohort from Kashgar, China. Independent prognostic indicators were identified using LASSO regularization combined with multivariate Cox regression. Seven predictive models were subsequently constructed, including CoxPH and six ML algorithms—RSF, GLMboost, decision tree, boosted tree(BT), DeepSurv, and NMTLR. Model discrimination and calibration were evaluated via C-index, time-dependent AUC, and integrated Brier score(iBS). Interpretability was elucidated through time-dependent permutation importance, partial dependence survival profiles, and multi-dimensional SurvSHAP(t) analysis.

Results

Ten predictors—age, Gleason score, M stage, PSA, surgery, marital status, histology, bone metastasis, N stage, and chemotherapy—were retained as independent factors. The boosted tree model achieved superior performance, with 1-, 3-, and 5-year OS AUCs exceeding 0.88 and iBS values below 0.160 across all cohorts. SurvSHAP(t) analysis ranked age, Gleason score, M stage, PSA, and marital status as the top five contributors. All core variables exhibited time-dependent increases in predictive weight. Notable antagonistic interactions were observed between age and M stage/Gleason score, while mild synergistic effects were found for age-chemotherapy and AJCC stage-marital status pairs.

Conclusions

The BT model demonstrates robust accuracy and interpretability for prostate cancer OS prediction. When combined with SurvSHAP(t) visualization, it may serve as a reliable auxiliary instrument for individualized diagnosis, treatment planning, and long-term surveillance.

1. Introduction

Prostate cancer represents the most frequently diagnosed urogenital malignancy in males, with its global incidence and mortality continuing to escalate at an alarming pace[1, 2]. In 2020, approximately 1.41 million new cases and 375,304 deaths were documented worldwide. Although Asian populations

exhibit substantially lower incidence rates compared with their Western counterparts, the disease burden in this region is projected to grow steadily through 2030[3–5]. Despite advances in early detection and targeted therapeutic strategies, long-term OS among affected patients remains suboptimal. Reliable survival estimation is therefore critical for tailoring treatment intensity and planning appropriate follow-up regimens[6].

Machine learning methodologies have demonstrated considerable promise in extracting meaningful patterns from complex clinical datasets. Nevertheless, the majority of existing prognostic models for prostate cancer suffer from inadequate interpretability, often functioning as "black boxes" that impede their translation into routine clinical practice[7, 8]. Explainable artificial intelligence(XAI) offers a pathway to overcome this barrier by rendering model decision processes transparent. However, interpretable prognostic tools specifically designed for prostate cancer—particularly those validated on independent Asian real-world cohorts—remain conspicuously scarce[9, 10].

This study developed multiple survival prediction models using large-scale SEER data and an independent multi-ethnic external cohort from Kashgar, Xinjiang. We systematically compared seven algorithms and comprehensively interpreted the optimal model via time-varying feature importance, partial dependence curves and SurvSHAP analysis. This work aims to provide an accurate, transparent and cross-population applicable prognostic tool for precise PCa treatment.

2. Materials and Methods

2.1 data sources.

This retrospective investigation included prostate cancer patients diagnosed between 2017 and 2021, identified from the SEER database using ICD-O-3 site code C61 and standard histological malignancy codes. Exclusion criteria encompassed missing OS information, age below 15 years, incomplete TNM staging, unknown histological grade, and absent summary stage data. Following these exclusions, 90,922 eligible SEER patients were randomly assigned in a 7:3 ratio to the training cohort(n = 63,645) and internal validation cohort(n = 27,277). The external validation set comprised 970 prostate cancer patients treated between 2016 and 2023 at a tertiary referral hospital in Kashgar, China.

All variables were independently extracted by two researchers and cross-verified to ensure data accuracy, with complete de-identification applied. The institutional review board granted a waiver of written informed consent for both datasets, given their retrospective and anonymized nature(Fig. 1).

2.2 variable preprocessing.

Fifteen clinical parameters with missing rates below 5% were retained for analysis, covering marital status, age, Gleason score, PSA level, TNM stage, histology, bone metastasis, and treatment modalities. For categorical variables, missing entries were imputed using the mode; for continuous age values,

mean imputation was applied to address sporadic missing data. All preprocessing procedures were consistently executed across the three cohorts using Python 3.12.9. Overall survival served as the primary endpoint throughout this study.

2.3 Prognostic factor screening.

All candidate variables were transformed into categorical formats and summarized as frequencies with percentages. To mitigate overfitting risks, the least absolute shrinkage and selection operator(LASSO) technique was first employed to identify the most relevant predictive features. Subsequently, multivariate Cox proportional hazards models were applied to determine independent prognostic factors, with corresponding 95% confidence intervals(CIs) computed for each retained variable.

2.4 Model development, tuning, and performance assessment

Seven survival models were fitted: CoxPH, GLMboost, RSF, BT, decision tree, NMTLR and DeepSurv. Hyperparameter optimization for all ML algorithms was conducted via 5-fold cross-validation combined with grid/random search on the training set, targeting the C-index, iBS, and time-dependent AUC as optimization criteria. Key tuned parameters for each model were recorded to ensure reproducibility.

Model discrimination was quantified using the C-index and time-dependent AUC, while calibration was evaluated through the integrated Brier score(iBS), where higher C-index/AUC values and lower iBS indicate superior predictive performance.

2.5 Interpretability analytics.

Three complementary approaches were adopted for model interpretation: (i)time-dependent permutation feature importance to trace dynamic variable contributions across follow-up; (ii)partial dependence survival profiles (PDPs) to visualize the isolated effect of each variable on predicted survival, adjusted for all other covariates; and (iii)aggregated SurvSHAP(t) analysis to quantify global feature importance, temporal weight trajectories, and pairwise interaction effects. The complete analytical pipeline is illustrated in Fig. 2.

2.6 Statistical analysis.

LASSO and Cox regression analyses were performed using SPSS 26.0. All machine learning modeling, performance evaluation, and interpretability visualization were carried out in Python 3.12.9 with Pandas, Scikit-survival, PyTorch, and SHAP libraries. Baseline characteristics across cohorts were compared using the Pearson chi-square test, with a two-sided $P < 0.05$ considered statistically significant.

3. RESULTS

3.1 Demographic and clinicopathological characteristics.

The distribution of most clinicopathological features was comparable across the training, internal validation, and external validation cohorts, with no statistically significant differences observed for the majority of variables(all $P > 0.05$). Notable exceptions included marital status ($P = 0.039$) and age distribution($P = 0.008$), which exhibited inter-cohort heterogeneity. These discrepancies are likely attributable to underlying demographic and regional variations between the predominantly Western SEER population and the multi-ethnic Northwest Chinese cohort served by the Kashgar center(Table 1).

Table 1
Baseline clinicopathological characteristics of prostate cancer patients across the training, internal validation, and external validation cohorts [n(%)]

Characteristic	**Training set (*n* = 63645) [n(%)]**	**Validation set (*n* = 27277) [n(%)]**	**External Validation set (*n* = 970) [n(%)]**	X^2	***P***
Grade				2.016	0.733
G1	10083(15.8)	4273(15.7)	147(15.2)		
G2	33842(53.2)	14595(53.5)	533(54.9)		
G3	19720(31.0)	8409(30.8)	290(29.9)		
Histology				4.943	0.764
Ad	63199(99.3)	27096(99.3)	966(99.6)		
ACN	153(0.2)	68(0.2)	2(0.2)		
DLN	160(0.3)	62(0.2)	1(0.1)		
EN	87(0.1)	27(0.1)	1(0.1)		
CMSN	46(0.1)	24(0.1)	0(0)		
Stage				1.825	0.768
Localized	45705(71.8)	19593(71.8)	692(71.3)		
Regional	14272(22.4)	6130(22.5)	213(22)		
Distant	3668(5.8)	1554(5.7)	65(6.7)		
AJCCstage				2.623	0.855
1	6206(9.8)	2634(9.7)	85(8.8)		
2	40419(63.5)	17384(63.7)	609(62.8)		
3	10607(16.7)	4521(16.6)	174(17.9)		
4	6413(10.1)	2738(10)	102(10.5)		
Tstage				5.761	0.450
T1	21250(33.4)	9005(33)	327(33.7)		
T2	28485(44.8)	12353(45.3)	418(43.1)		
T3	12988(20.4)	5558(20.4)	210(21.6)		
T4	922(1.4)	361(1.3)	15(1.5)		

Characteristic	Training set (*n* = 63645) [n(%)]	Validation set (*n* = 27277) [n(%)]	External Validation set (*n* = 970) [n(%)]	X^2	*P*
Nstage				0.673	0.714
N0	59656(93.7)	25569(93.7)	903(93.1)		
N1	3989(6.3)	1708(6.3)	67(6.9)		
Mstage				0.937	0.480
M0	60020(94.3)	25733(94.3)	908(93.6)		
M1	3625(5.7)	1544(5.7)	62(6.4)		
Surgery				4.389	0.111
No	28778(45.2)	12132(44.5)	430(44.3)		
Yes	34867(54.8)	15145(55.5)	540(55.7)		
Radiothearepy				4.671	0.587
No	35628(56)	15398(56.5)	561(57.8)		
Beam radiation	22134(34.8)	9389(34.4)	313(32.3)		
Radioactive implants	3294(5.2)	1396(5.1)	51(5.3)		
Combination of beam with implants or isotopes	2589(4.1)	1094(4)	45(4.6)		
Chemotherapy				4.294	0.117
No	62356(98.0)	26781(98.2)	950(97.9)		
Yes	1289(2.0)	496(1.8)	20(2.1)		
Bone Metastasis				0.062	0.970
No	60465(95)	25918(95)	920(94.8)		
Yes	3180(5)	1359(5)	50(5.2)		
Marital				13.254	0.039
Married	48094(75.6)	20756(76.1)	704(72.6)		
Single	7805(12.3)	3262(12)	126(13)		
Separated	5286(8.3)	2234(8.2)	106(10.9)		
Widowed	2460(3.9)	1025(3.8)	34(3.5)		

Characteristic	Training set (*n* = 63645) [n(%)]	Validation set (*n* = 27277) [n(%)]	External Validation set (*n* = 970) [n(%)]	X^2	*P*
PSA(ng/ml)				6.550	0.364
< 4	6276(9.9)	2642(9.7)	102(10.5)		
4.1 ~ 10	36531(57.4)	15867(58.2)	540(55.7)		
10.1 ~ 20	11610(18.2)	4902(18)	187(19.3)		
> 20	9228(14.5)	3866(14.2)	141(14.5)		
Age(year)				17.397	0.008
< 60	14984(23.5)	6370(23.4)	258(26.6)		
61 ~ 70	28700(45.1)	12426(45.6)	418(43.1)		
71 ~ 80	16453(25.9)	7096(26)	230(23.7)		
> 81	3508(5.5)	1385(5.1)	64(6.6)		
Gleason score				0.597	0.963
≤ 6	14147(22.2)	6101(22.4)	222(22.9)		
7	31108(48.9)	13335(48.9)	474(48.9)		
≥ 8	18390(28.9)	7841(28.7)	274(28.2)		

3.2 Analysis of prognostic factors of prostate cancer.

From an initial pool of 15 clinical parameters, LASSO regression with a significance threshold of $P < 0.05$ identified 10 variables that were retained for further analysis(Table 2, Fig. 3). These selected predictors were subsequently entered into a multivariate Cox proportional hazards model. The results confirmed that all 10 variables-namely Gleason Score, Age, PSA, M stage, Marital Status, Surgery, Bone Metastasis, Histology, Chemotherapy, and N stage-achieved statistical significance($P < 0.05$), thereby establishing them as independent risk factors for overall survival in prostate cancer(Table 3).

Table 2
Variables retained via LASSO regression with corresponding coefficients and significance levels

Variable	Coefficient	Z_Score	P_Value
Age	0.464	45.389	< 0.001
GleasonScore	0.234	17.803	< 0.001
Mstage	0.201	10.811	< 0.001
PSA	0.189	18.125	< 0.001
Surgery	-0.182	-11.711	< 0.001
Maritalstatus	0.150	16.897	< 0.001
Histology	-0.092	-7.077	< 0.001
BoneMets	0.091	5.651	< 0.001
Nstage	0.065	8.897	< 0.001
Chemotherapy	-0.038	-3.462	< 0.001
Radiotherapy	0.023	1.535	0.123
Tstage	0.019	1.114	0.289
AJCCstage	0	0.122	0.903
Grade	0	0	1.000
Stage	0	0.158	0.875

Table 3
Univariate and multivariate Cox regression analyses of independent prognostic factors associated with overall survival in prostate cancer

Characteristic	Univariate analysis		Multivariate analysis	
	***HR* (95%*CI*)**	***P***	***HR* (95%*CI*)**	***P***
Histology				
Ad	Reference value		Reference value	
ACN	0.489(0.284,0.842)	0.010	0.798(0.460,1.365)	0.401
DLN	1.601(1.182,2.168)	0.002	1.307(0.965,1.771)	0.084
EN	3.702(2.734,5.012)	< 0.001	2.585(1.908,3.503)	< 0.001
CMSN	0.591(0.246,1.420)	0.240	0.743(0.309,1.786)	0.507
Nstage				
N0	Reference value		Reference value	
N1	3.312(3.143,3.490)	< 0.001	1.306(1.230,1.388)	< 0.001
Mstage				
M0	Reference value		Reference value	
M1	9.416(9.003,9.849)	< 0.001	2.252(1.962,2.584)	< 0.001
Surgery				
No	Reference value		Reference value	
Yes	0.382(0.367,0.398)	< 0.001	0.747(0.715,0.780)	< 0.001
Chemotherapy				
No	Reference value		Reference value	
Yes	5.561(5.169,5.983)	< 0.001	1.172(1.078,1.273)	< 0.001
Bone Metastasis				
No	Reference value		Reference value	
Yes	9.878(9.428,10.349)	< 0.001	1.573(1.374,1.802)	< 0.001
Marital				
Married	Reference value		Reference value	
Single	1.484(1.407,1.566)	< 0.001	1.566(1.483,1.653)	< 0.001
Separated	1.491(1.401,1.587)	< 0.001	1.426(1.339,1.518)	< 0.001

Characteristic	Univariate analysis		Multivariate analysis	
	HR (95%*CI*)	*P*	*HR* (95%*CI*)	*P*
Widowed	3.025(2.827,3.237)	< 0.001	1.628(1.519,1.746)	< 0.001
PSA(ng/ml)				
< 4	Reference value		Reference value	
4.1 ~ 10	0.799(0.742,0.861)	< 0.001	0.758(0.704,0.817)	< 0.001
10.1 ~ 20	1.572(1.454,1.700)	< 0.001	1.043(0.963,1.129)	0.306
> 20	4.046(3.757,4.358)	< 0.001	1.420(1.310,1.541)	< 0.001
Age(year)				
< 60	Reference value		Reference value	
61 ~ 70	1.691(1.584,1.804)	< 0.001	1.612(1.510,1.721)	< 0.001
71 ~ 80	3.408(3.196,3.635)	< 0.001	2.757(2.577,2.948)	< 0.001
> 81	10.854(10.105,11.659)	< 0.001	5.916(5.474,6.394)	< 0.001
Gleason score				
≤ 6	Reference value		Reference value	
7	1.283(1.207,1.365)	< 0.001	1.062(0.997,1.130)	0.061
≥ 8	4.087(3.854,4.333)	< 0.001	1.736(1.627,1.851)	< 0.001

3.3 Model comparison.

All seven constructed models demonstrated acceptable levels of discrimination and calibration, with integrated Brier scores consistently below 0.08 across all evaluated datasets. Among these, the boosted tree algorithm consistently outperformed its counterparts across all performance metrics. Specifically, the BT model achieved the following results: in the training set, a C-index of 0.785, an integrated AUC of 0.823, and an iBS of 0.071; in the internal validation set, a C-index of 0.778, an integrated AUC of 0.825, and an iBS of 0.071; and in the external Chinese cohort, a C-index of 0.793, an integrated AUC of 0.823, and an iBS of 0.069(Table 4). The time-dependent ROC curves(Fig. 4) further substantiated the model's robust predictive capacity. At 1-, 3-, and 5-year time points, the AUC values for the training set were 0.831, 0.825, and 0.813, respectively; for the internal validation set, corresponding figures were 0.845, 0.819, and 0.810; and for the external validation set, the AUCs were 0.833, 0.830, and 0.805, respectively.

Table 4
Comparative predictive performance of seven survival models across the training, internal validation, and external validation cohorts

Training set	C-index	Integrated C/D AUC	Integrated brier score
Coxph	0.775	0.784	0.073
RSF	0.779	0.787	0.072
GLMboost	0.766	0.775	0.076
Boosting trees	0.785	0.823	0.071
Decision trees	0.766	0.774	0.072
DeepSurv	0.783	0.793	0.071
NMTLR	0.785	0.794	0.061
Validation set			
Coxph	0.772	0.775	0.072
RSF	0.774	0.776	0.071
GLMboost	0.762	0.765	0.075
Boosting trees	0.778	0.825	0.071
Decision trees	0.757	0.758	0.073
DeepSurv	0.776	0.780	0.071
NMTLR	0.777	0.779	0.062
External Validation set			
Coxph	0.781	0.797	0.072
RSF	0.791	0.805	0.069
GLMboost	0.774	0.785	0.077
Boosting trees	0.793	0.823	0.069
Decision trees	0.782	0.780	0.073
DeepSurv	0.798	0.813	0.068
NMTLR	0.795	0.810	0.062

3.4 Model Interpretability.

3.4.1 Time-dependent feature importance

Time-dependent permutation feature importance was evaluated through two complementary metrics: the increment in Brier score and the decrement in time-dependent C/D AUC(Fig. 5). These two measures revealed notably distinct temporal patterns. When assessed by Brier score loss, M stage emerged as the predominant predictor during the early to mid-follow-up period, specifically for survival times less than approximately 54 months. Beyond this threshold, Age superseded M stage as the variable exerting the greatest influence on predictive performance. In contrast, when evaluated via C/D AUC loss, Age consistently ranked as the most critical feature across the entire observation window of 0 to 70 months, demonstrating sustained and stable contribution to survival discrimination throughout follow-up. These divergent temporal trajectories underscore the dynamic heterogeneity inherent in prognostic factors for prostate cancer, providing compelling evidence for adopting time-stratified risk evaluation and adaptive clinical surveillance strategies in routine practice.

3.4.2 partial dependence survival profiles

Partial dependence survival profiles offer a global explanation of the BT model's predictive behavior(Fig. 6). These plots illustrate how the predicted survival probability-defined as the model-estimated likelihood of a patient surviving beyond a given time point-varies in response to changes in a single predictor, while all other covariates are held constant at their observed values in the training dataset. In Fig. 6, narrow and closely overlapping confidence bands indicate that variations in the corresponding variable have minimal impact on the overall predicted survival trajectory. Conversely, wider and non-overlapping bands suggest that even modest alterations in that variable can produce substantial shifts in predicted survival probability. Using this interpretive framework, we observed that modifications in variables such as Age, M stage, Gleason score, Chemotherapy, Marital status, PSA, Histology, and Bone metastases all led to marked changes in predicted survival. Notably, patients aged 80 years or older exhibited a considerably steeper decline in predicted survival compared with those under 60 years of age. Similar patterns were evident among individuals presenting with M1-stage disease, Gleason scores exceeding 8, receipt of chemotherapy, widowed marital status, PSA levels above 20ng/mL, En-type histology, or documented bone metastasis.

3.4.3 Aggregated SurvSHAP values summary

Global importance ranking, depicted in Fig. 7A, placed Age first, followed in descending order by Gleason score, M stage, PSA, and Marital status. Figure 7B illustrates the temporal variability of each variable's significance, with the y-axis representing the average absolute SurvSHAP(t) value at each time point, highlighting how feature importance fluctuates over the follow-up period. The bee swarm plot in Fig. 7C provides a detailed view of individual patient-level contributions: variables are arranged by their mean absolute SurvSHAP(t) values in descending order, with each point representing a single patient's SurvSHAP(t) value for a specific variable; the x-axis indicates the magnitude and direction of the

feature's impact on OS prediction (positive values denote increased mortality risk, while negative values signify reduced risk), and the color gradient encodes the original feature values.

Temporal trajectories of the top five predictors over the 0 to 70-month follow-up period are presented in Fig. 7D. All five variables demonstrated a consistent upward trend in absolute SurvSHAP(t) values as follow-up lengthened, indicating that their prognostic influence intensifies over time. Among these, Age exhibited the most pronounced increase in contribution, with its SurvSHAP(t) value rising continuously and reaching the highest magnitude by 70 months-a finding that underscores its growing importance in long-term survival stratification. Gleason score and M stage also displayed steady, moderate increases throughout follow-up, with both factors exerting meaningful impact on both short- and long-term outcomes. PSA and Marital status showed slower yet persistent growth, with their prognostic relevance becoming more evident during the mid-to-late follow-up phases. These trajectories are consistent with the permutation-based importance findings, further corroborating the dynamic nature of prognostic factors in prostate cancer.

Two-way interaction analyses were conducted using SHAP plots to evaluate potential effect modifications among key prognostic features(Fig. 7E), with correlation coefficients and associated p-values calculated to identify statistically meaningful interactions. Among all evaluated pairs, four reached statistical significance. The Age-M stage pair showed a moderate negative correlation ($r=-0.396$, $p < 0.0001$), indicating a significant antagonistic interaction: the adverse prognostic impact of metastatic disease is modified by age, with older patients demonstrating a weaker independent effect of M stage on survival, whereas younger patients with metastatic disease experience a more pronounced risk gradient. The Age-Gleason score pair exhibited a weak negative correlation ($r=-0.177$, $p = 0.0021$), reflecting a mild but significant interaction wherein both factors operate largely independently in survival stratification. The Age-Chemotherapy pair demonstrated a moderate positive correlation ($r = 0.256$, $p < 0.0001$), revealing that the survival benefit associated with chemotherapy varies distinctly across different age groups. The AJCC stage-Marital status pair showed a mild positive correlation ($r = 0.123$, $p = 0.0334$), demonstrating that unfavorable marital status-such as being widowed-may exacerbate the adverse prognostic effect of advanced tumor stage. Other evaluated pairs, including Age-Marital status ($r=-0.013$, $p = 0.8173$) and Chemotherapy-Marital status ($r = 0.087$, $p = 0.1306$), exhibited no significant associations. Collectively, these findings underscore the clinical relevance of jointly considering age with metastatic burden, treatment response, and social support when individualizing surveillance and therapeutic regimens for prostate cancer patients.

4. DISCUSSION

This study constructed seven predictive models combining traditional Cox regression and six machine learning algorithms based on large-scale SEER prostate cancer data and an independent real-world Kashgar Chinese external cohort. Comprehensive performance comparison and multi-dimensional interpretable analysis via permutation-based time-dependent feature importance, partial dependence survival curves, and aggregated SurvSHAP quantification were systematically conducted to identify the

optimal boosting trees(bt) model for overall survival(OS) prediction of prostate cancer. The multi-cohort verification and transparent predictive mechanism interpretation jointly provide an accurate, generalizable, and clinically understandable risk stratification tool for prostate cancer individualized management.

4.1 Baseline Cohort Characteristics and Cohort Balance Analysis

We first compared the clinicopathological characteristics of the SEER training set, internal validation set and Kashgar external validation cohort. Pearson chi-square test results showed most indicators were well-balanced across three groups, while statistically significant differences existed in marital status and age distribution. Such discrepancies can be reasonably attributed to racial, regional, and healthcare demographic disparities between Western populations in the SEER database and multi-ethnic residents in Xinjiang, Northwest China[11, 12]. Chinese prostate cancer patients generally present with older median age at diagnosis and distinct marital composition compared with Western counterparts, consistent with existing epidemiological reports on Asian prostate cancer populations[13]. Despite baseline heterogeneity in age and marital status, the subsequent model evaluation still maintained stable discrimination and calibration on the external cohort, strongly proving the boosting trees model`s cross-population generalizability, rather than being limited to Western patient datasets.

4.2 Screening and Clinical Rationality of Independent Prognostic Factors

LASSO regression combined with multivariate Cox proportional hazards regression screened out 10 independent prognostic factors for prostate cancer overall survival, namely Age, GleasonScore, Mstage, PSA, Surgery, Maritalstatus, Histology, BoneMets, Nstage and Chemotherapy. Among them, Age, Gleason score and Mstage showed remarkably high hazard ratios in multivariate Cox analysis, standing as the most powerful adverse prognostic markers consistent with clinical consensus: advanced age impairs treatment tolerance, high Gleason score indicates highly invasive undifferentiated lesions, and distant metastasis represents irreversible advanced disease[14–16]. Bone metastasis, lymph node involvement and elevated PSA also predicted worse survival, while surgery brought survival benefits[17]. Unmarried patients and those with aggressive histological subtypes faced higher mortality, possibly due to insufficient social support and more malignant biological behavior. These multi-dimensional demographic, pathological and therapeutic predictors allow our model to realize more comprehensive risk assessment than traditional single-index scoring tools.

4.3 Superiority of the Boosting Trees Model in Predictive Performance

C-index, integrated C/D AUC and integrated Brier score were adopted to comprehensively compare seven predictive models across training, internal validation and external cohorts. The boosting trees model

achieved the best performance in all four evaluation metrics: 1-, 3-, 5-year OS AUC all exceeded 0.80 in all datasets, integrated Brier score remained below 0.07. The outstanding performance of the bt model originates from its iterative ensemble framework. Unlike Cox regression constrained by the proportional hazards assumption and linear correlation limitation, boosting trees iteratively trains weak learners to capture complex nonlinear relationships and high-order interactions among multi-clinical variables (e.g., the joint effect of age and distant metastasis)[18, 19]. Meanwhile, built-in regularization avoids overfitting on large SEER training data. Other machine learning models including RSF, GLM, DeepSurv and NMTLR displayed decent predictive capacity but lagged behind boosting trees in discrimination and calibration. Simple decision tree had the weakest performance due to limited feature interaction capturing ability. Crucially, the model maintained stable performance on the independent Kashgar Chinese external cohort without obvious performance attenuation. Most existing prostate cancer prognostic models are only internally verified on Western databases and lack external validation on Asian populations. Our cross-regional, cross-ethnic verification result fills this research gap, demonstrating this model can be reliably applied to clinical risk assessment of Chinese prostate cancer patients.

4.4 Time-Dependent Dynamic Changes of Prognostic Feature Importance

Permutation-based time-dependent feature importance revealed obvious temporal heterogeneity of prognostic factors, laying a foundation for time-stratified clinical management[20]. Two evaluation indicators yielded inconsistent trends: according to Brier score loss, Mstage exerted the greatest predictive impact when survival time was less than approximately 54 months, whereas Age became the dominant predictor after 54 months. However, the C/D AUC loss metric indicated Age retained the highest predictive weight across the whole 0–70 month follow-up period. Such divergence arises because AUC reflects the model`s ability to distinguish high- and low-risk patients, while Brier score evaluates overall predictive calibration accuracy. Age consistently differentiates patient risks at all time points, yet distant metastasis status governs the accuracy of short- and mid-term survival prediction, while age drives long-term predictive calibration. Clinically, this dynamic pattern reminds clinicians to prioritize surveillance of metastatic lesions for patients with short-to-medium predicted survival, and carry out age-adapted supportive and anti-tumor interventions for long-term survivors. Conventional static risk scores fail to capture such time-varying feature weights, which demonstrates the clinical necessity of time-dependent interpretable machine learning models.

4.5 Independent Feature Effects Reflected by Partial Dependence Survival Plots

Partial dependence survival profiles (PDPs) visually quantified the independent survival impact of each single variable after adjusting all other covariates, further verifying the rationality of the 10 prognostic factors screened in this study[21, 22]. Patients with bone metastasis, M1 distant metastasis, high Gleason score(≥ 8), high histological grade, elevated PSA(> 20ng/mL) and age ≥ 80 years displayed rapidly declining survival curves; conversely, patients receiving radical surgery presented significantly higher

long-term survival probability. Widowed patients also showed obviously worse survival compared with married groups. These visualized curves translate abstract statistical hazard ratios into intuitive survival trend graphs, which are more acceptable and operable for frontline urologists without statistical background. When facing an individual patient, clinicians can rapidly judge the survival trend brought by a single abnormal indicator via PDP reference, and combine multiple adverse factors to comprehensively judge overall risk, realizing visual shared clinical decision-making between doctors and patients.

4.6 Quantitative Global and Temporal Feature Contribution via SurvSHAP Analysis

Aggregated SurvSHAP analysis provided quantitative global, time-varying and interactive interpretation of the boosting trees model, making the “black-box” prediction logic fully traceable[23, 24]. Global SurvSHAP mean absolute value ranking showed Age, Gleason score, Mstage, PSA and Marital status were the top five prognostic contributors, consistent with LASSO and time-dependent importance results, mutually validating the core status of metastatic and age-related indicators. Temporal SurvSHAP trajectories revealed the absolute contribution of all key predictors gradually increased with prolonged follow-up, meaning these variables exert more prominent predictive effects on long-term survival rather than short-term prognosis. Two-way SurvSHAP interaction plots further identified four statistically significant feature pairwise correlations:

Age vs. Mstage: moderate negative antagonistic interaction ($r=-0.396$, $p < 0.0001$). The adverse survival risk caused by distant metastasis was weakened in elderly patients, while younger metastatic prostate cancer patients suffered a far more severe risk deterioration, indicating young metastatic patients need more intensive anti-tumor intervention;

Age vs. Gleason score: weak negative correlation ($r=-0.177$, $p = 0.0021$). Gleason score and age independently affect survival with little mutual modification effect;

Age vs. Chemotherapy: moderate positive correlation ($r = 0.256$, $p < 0.0001$), revealing chemotherapy survival benefits differ significantly across age groups, requiring age-adapted chemotherapy intensity;

AJCCStage vs. Marital status: mild positive synergistic correlation ($r = 0.123$, $p = 0.0334$). Widowed patients with advanced tumor stage carry superimposed poor prognosis, necessitating additional psychological and life support besides anti-tumor therapy. Other feature pairs exhibited non-significant weak correlations, indicating they independently predict survival without synergistic or antagonistic modification effects. All interaction results were stabilized by 10-fold cross-validation and bootstrap resampling, ensuring statistical robustness. Traditional Cox models only output independent hazard ratios and cannot quantify feature interactions, while SurvSHAP supplements this critical dimension. The interaction information helps clinicians identify high-risk subgroups requiring customized combined therapy and supportive management, forming multi-angle individualized treatment evidence.

4.7 Advantages Compared with Existing Prostate Cancer Prognostic Models

Most previous prostate cancer prediction tools have two prominent defects: insufficient external cross-population verification and lack of multi-dimensional interpretability. Traditional Cox nomograms only provide static point risk values without disclosing how predictors change their impact over follow-up time, nor can they quantify feature interactions. Conventional machine learning models only output prediction results without interpretable evidence, limiting clinical transformation due to the “black-box” defect. In contrast, this study’s boosting trees model integrates three complementary interpretive systems(time-dependent permutation importance, partial dependence curves, SurvSHAP global/temporal/interactive quantification). It not only achieves higher discrimination and calibration than Cox, RSF, DeepSurv and other algorithms, but also realizes transparent, quantitative, visual predictive logic explanation. Moreover, the model passed external verification on Chinese multi-ethnic cohort from Kashgar, filling the gap of Western database models lacking Asian population applicability. The multi-angle interpretable results can directly guide staged follow-up and subgroup individualized treatment, possessing stronger clinical operability than existing predictive tools[25].

4.8 Study Limitations

Several limitations of this research need to be acknowledged. First, the SEER database lacks detailed data on molecular biomarkers(AR-V7, PSMA), specific chemotherapy/hormonal medication regimens and patient comorbidity burden, which cannot be incorporated into feature screening and may partially limit model predictive potential. Second, the external validation cohort was single-center data from Kashgar with a relatively small sample size; multi-center cohorts covering different regions of China are required to further verify universal applicability in Chinese prostate cancer populations. Third, this is a retrospective observational study, inherent selection bias from database recording and unmeasured confounding factors cannot be completely excluded. Finally, the model only predicts overall survival, without establishing separate prediction frameworks for cancer-specific survival and biochemical recurrence.

4.9 Future Research Directions

Future work can be expanded in four directions:

1)Integrate molecular markers, imaging radiomics and comorbidity indicators to enrich feature sets and further optimize model accuracy;
2)Collect multi-center Chinese prostate cancer cohorts nationwide to conduct large-sample external validation of the bt model;

3)Develop visual web-based/miniapp clinical decision support tools embedded with all interpretability charts(time importance, PDP, SurvSHAP) for convenient real-time clinical risk calculation;

4)Extend model endpoints to biochemical recurrence and cancer-specific survival, building multi-outcome integrated interpretable prediction systems to meet diversified clinical assessment demands.

4.10 Brief Summary of Clinical Significance

In summary, this study constructed and externally validated an interpretable boosting trees OS prediction model for prostate cancer based on Western SEER data and Chinese real-world cohort. Multi-dimensional interpretive analysis via time-dependent feature importance, partial dependence curves and SurvSHAP quantification clarified the static, temporal and interactive prognostic effects of all core predictors[26, 27]. The model has superior predictive performance and cross-population generalizability, and all predictive mechanisms can be quantitatively and visually explained, overcoming the black-box limitation of traditional machine learning. It can serve as a reliable auxiliary tool for urologists to conduct dynamic risk stratification, staged surveillance planning and individualized treatment formulation for prostate cancer patients, promoting the development of precision prostate cancer management.

5. Conclusion

In this study, we constructed and externally validated an interpretable boosting trees model for predicting overall survival in prostate cancer patients using a large-scale SEER database cohort and a real-world multi-ethnic Chinese cohort from Kashgar, Xinjiang. After systematic comparison of seven predictive algorithms, the boosting trees model achieved optimal discrimination and calibration performance, with 1-, 3-, and 5-year AUC values exceeding 0.80 and stable low integrated Brier scores across training, internal validation and external validation sets. LASSO and multivariate Cox regression identified 10 independent prognostic factors covering age, tumor pathology, metastasis status and treatment modalities, among which Age, GleasonScore and Mstage exhibited the strongest adverse prognostic effects. Multi-layered interpretive analyses further uncovered time-varying prognostic contributions of key variables and meaningful feature interactions. Permutation-based time-dependent importance showed Mstage dominated predictive calibration for survival shorter than 54 months, while Age became the primary predictor beyond 54 months; meanwhile, Age maintained the core discriminatory effect throughout the entire 0-70-month follow-up period. SurvSHAP quantification visualized the global ranking, temporal trajectories and pairwise interaction effects of all predictors, revealing age-dependent differences in chemotherapy efficacy and synergistic risks between advanced AJCC stage and disadvantaged marital status. Collectively, this high-performance, fully interpretable model realizes dynamic, individualized risk stratification and provides clear clinical guidance for time-staged surveillance and tailored therapy. It possesses good generalizability for both Western and Northwest Chinese prostate cancer populations and can serve as an intuitive auxiliary prognostic tool to support clinicians in precision diagnosis and long-term disease management.

Declarations

Competing Interests: All the authors of this paper declare that there is no financial or non - financial interests that are directly or indirectly related to the work submitted for publication. This ensures the objectivity and integrity of our research presented in this paper.

Ethical Statement: This study was conducted in accordance with the Declaration of Helsinki.The training and validation sets of prostate cancer prognostic factor data used in this study are free and publicly available. Therefore, the SEER database data in this study is exempt from review and informed consent requirements. The prognostic factors data for prostate cancer in the external validation set (First People's Hospital of Kashi) were approved by the Ethics Committee of Kashi First People's Hospital(Ethics No.2024KJB No.04) and exempted from the requirement of informed consent (retrospective study).

Funding Statement

This work was supported by The second group of the Tianshan Talent Training Program: The Youth Support Talent Project [grant number 2023TSYCQNTJ0009];National Natural Science Foundation of China Project [grant number 82360359]; Special Fund Project for Central Government-Guided Local Science and Technology Development [grant number 2024B0202010005, XW];National Science and Technology Innovation 2030- Major Project of "New Generation Artificial Intelligence"[grant number 2022ZD0160700];Robust, Fair, and Trustworthy Comprehensive Evaluation System for Models [grant number 2022ZD0160705]; Xinjiang Key Laboratory of Artificial intelligence Assisted Imaging Diagnosis Fund [grant number XJRGZN2024011].

Author Contribution

S.A. and A.T. conceived the study, extracted SEER data, developed initial ML models, and drafted the manuscript.S.A. and P.W. collected Chinese external validation data, assisted LASSO/Cox analyses, and revised the manuscript.S.A. and J.B. conducted interpretability analyses (time-dependent importance, partial survival plots) and prepared supplementary tables.S.A. and G.W. and A.T. performed data quality control/cleaning and assisted performance evaluations (C-index, AUC, Brier score).S.A. and M.Y. implemented RSF/DeepSurv models, validated model results, and summarized performance metrics.S.A. and Y.Q. supervised the study, provided clinical insights, and revised the manuscript.S.A. and A.T. oversaw statistical work, guided SurvSHAP analysis, and finalized the manuscript.All authors reviewed the manuscript.

References

1. Lowrance W, Dreicer R, Jarrard DF, et al. Updates to Advanced Prostate Cancer: AUA/SUO Guideline (2023). J Urol. 2023;209(6):1082–90.
2. Kumar Am S, Rajan P, Alkhamees M, et al. Prostate cancer theragnostics biomarkers: An update. Investig Clin Urol. 2024;65(6):527–39.

3. 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 Cancer J Clin. 2021;71(3):209–49.
4. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024 May-Jun;74(3):229–63.
5. Vaccarella S, Li M, Bray F, et al. Prostate cancer incidence and mortality in Europe and implications for screening activities: population based study. BMJ. 2024;386:e077738.
6. Sekhoacha M, Riet K, Motloung P, et al. Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches. Molecules. 2022;27(17):5730.
7. Theodore Armand TP, Nfor KA, Kim JI, et al. Applications of Artificial Intelligence, Machine Learning, and Deep Learning in Nutrition: A Systematic Review. Nutrients. 2024;16(7):1073.
8. Chen G, Dai X, Zhang M, et al. Machine learning-based prediction model and visual interpretation for prostate cancer. BMC Urol. 2023;23(1):164.
9. Guan C, Gong A, Zhao Y, et al. Interpretable machine learning model for new-onset atrial fibrillation prediction in critically ill patients: a multi-center study. Crit Care. 2024;28(1):349.
10. Qi X, Wang S, Fang C, et al. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox Biol. 2025;79:103470.
11. Chen L, Xu YX, Wang YS, et al. Integrative Chinese-Western medicine strategy to overcome docetaxel resistance in prostate cancer. J Ethnopharmacol. 2024;331:118265.
12. Li S, Jing S, Zhou J, et al. Chinese medicines as therapeutic options for treating prostate cancer: Therapeutic effects and underlying mechanisms. J Ethnopharmacol. 2026;354:120471.
13. Cheng B, Luo T, Wu Y, et al. Urinary exosomal FAM153C-RPL19 chimeric RNA as a diagnostic and prognostic biomarker for prostate cancer in Chinese patients. Cancer Lett. 2025;631:217938.
14. Mohammadi S, Imani B, Saeedi S, et al. Identification of Factors Affecting Prostate Cancer Using Machine Learning Methods: A Systematic Review. Asian Pac J Cancer Prev. 2025;26(5):1519–28.
15. Tang S, Zhang H, Liang J, et al. Prostate cancer treatment recommendation study based on machine learning and SHAP interpreter. Cancer Sci. 2024;115(11):3755–66.
16. Mohammed Nawi A, Masdor NA, Othman R, et al. Survival Rate and Prognostic Factors of Localised Prostate Cancer in Southeast Asian Countries: A Systematic Review with Meta-Analysis. Asian Pac J Cancer Prev. 2023;24(9):2941–2095.
17. Higashiyama S, Yoshida A, Kawabe J. Predicting the Prognosis of Prostate Cancer Bone Metastasis Using the Bone Scan Index and Hot Spots Calculated Using VSBONE® Bone Scan Index from Tc-99m-Hydroxymethylene Diphosphonate Bone Scintigraphy. Urol Int. 2022;106(9):963–9.
18. Molaei S, Moazen H, Niazkar HR, et al. Application of boosted trees to the prognosis prediction of COVID-19. Health Sci Rep. 2024;7(5):e2104.

19. Godoy LC, Farkouh ME, Austin PC, et al. Predicting left main stenosis in stable ischemic heart disease using logistic regression and boosted trees. Am Heart J. 2023;256:117–27.
20. Zhu J, Chen H, Duan S, et al. Development and validation of a simplified time-dependent interpretable machine learning-based survival model for older adults with multimorbidity. NPJ Aging. 2025;12(1):10.
21. Li M, Han S, Liang F, et al. Machine Learning for Predicting Risk and Prognosis of Acute Kidney Disease in Critically Ill Elderly Patients During Hospitalization: Internet-Based and Interpretable Model Study. J Med Internet Res. 2024;26:e51354.
22. Kallah-Dagadu G, Mohammed M, Nasejje JB, et al. Breast cancer prediction based on gene expression data using interpretable machine learning techniques. Sci Rep. 2025;15(1):7594.
23. Gong Z, Zhou L, He Y, et al. Explainable Machine Learning Model for Predicting Postoperative Survival in Patients With Locally Advanced Gastric Cancer. Cancer Med. 2025;14(22):e71408.
24. Dai H, Yu Z, Zhao Y, et al. Integrating machine learning models with multi-omics analysis to decipher the prognostic significance of mitotic catastrophe heterogeneity in bladder cancer. Biol Direct. 2025;20(1):56.
25. Merseburger AS, Agarwal N, Bhaumik A, et al. Apalutamide plus androgen deprivation therapy in clinical subgroups of patients with metastatic castration-sensitive prostate cancer: A subgroup analysis of the randomised clinical TITAN study. Eur J Cancer. 2023;193:113290.
26. Luo Q, Zhang Q, Liu H, et al. Time-dependent interpretable survival prediction model for second primary NSCLC patients. Int J Med Inf. 2025;195:105771.
27. Hajipour Khire Masjidi B, Bahmani S, Sharifi F et al. CT-ML: Diagnosis of Breast Cancer Based on Ultrasound Images and Time-Dependent Feature Extraction Methods Using Contourlet Transformation and Machine Learning. Comput Intell Neurosci. 2022;2022:1493847. 10.1155/2022/1493847. Retraction in: Comput Intell Neurosci. 2023;2023:9814037.

Figures

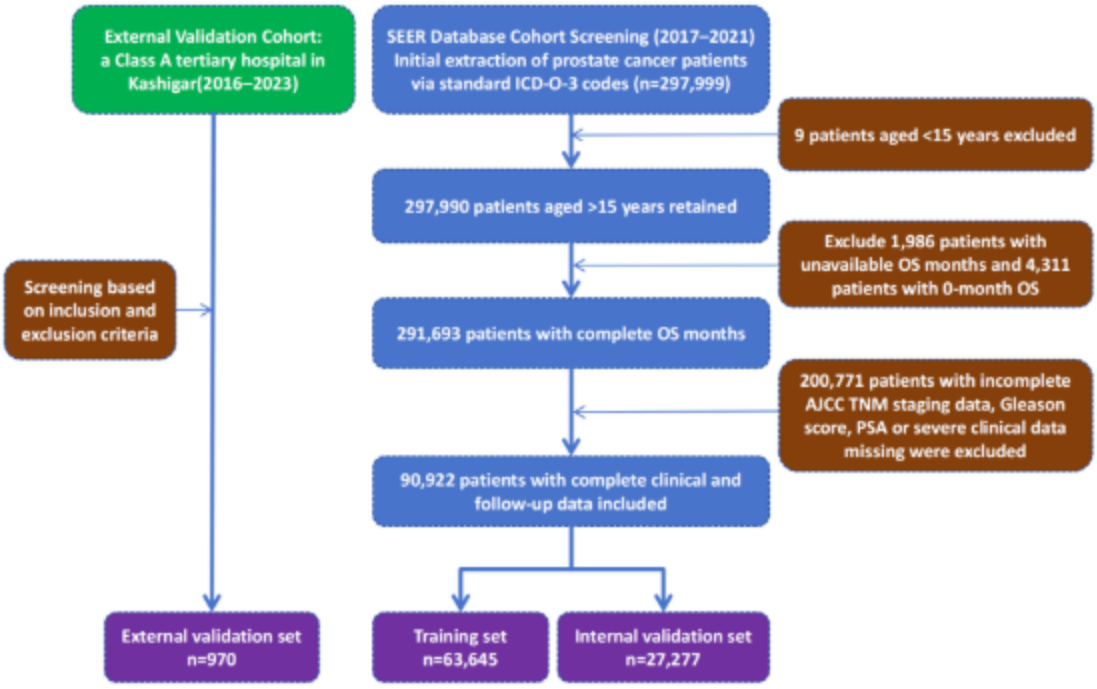


Figure 1

Flow diagram depicting the patient selection and cohort assembly process for prostate cancer cases

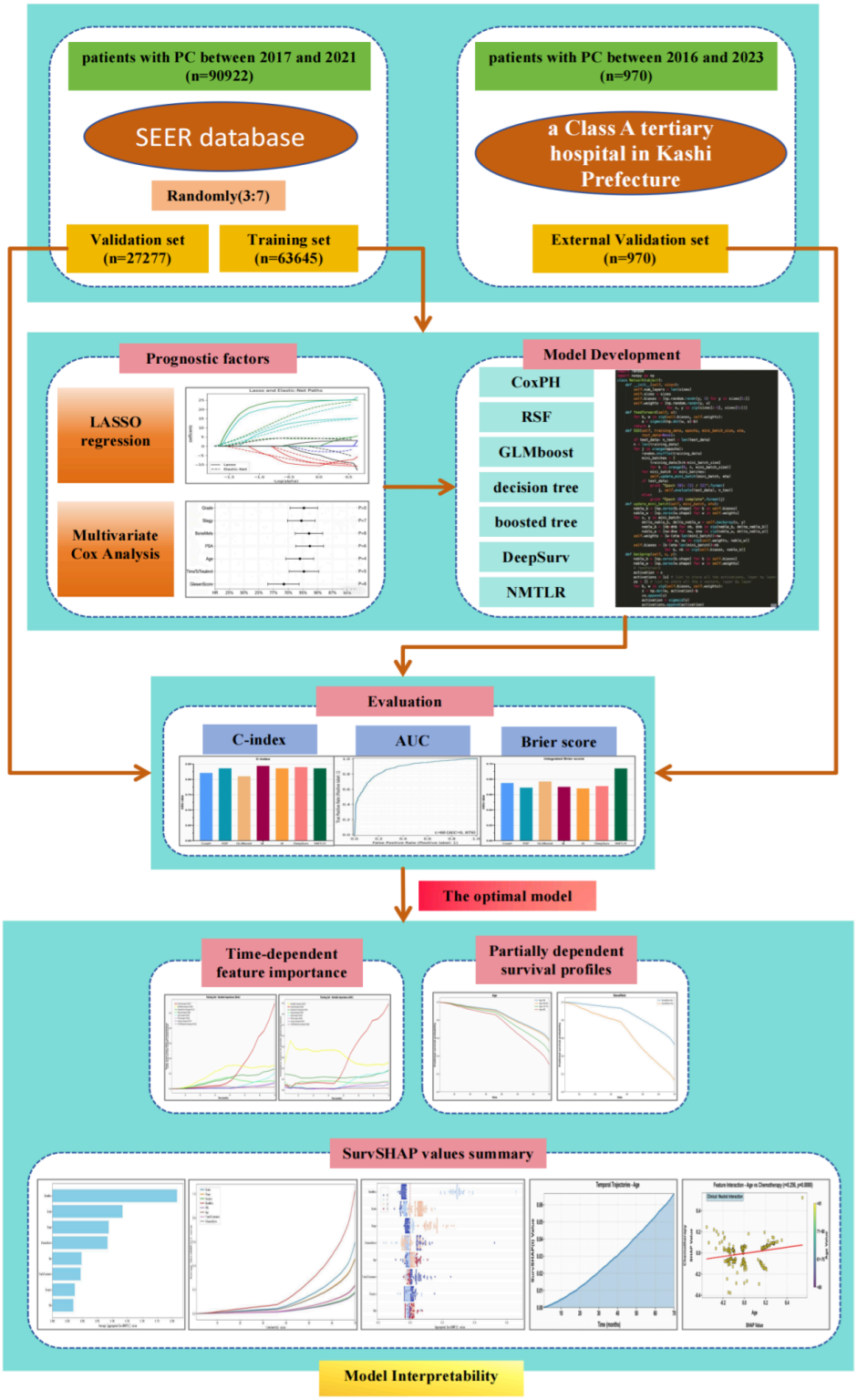


Figure 2

Schematic representation of the overall study design and analytical workflow

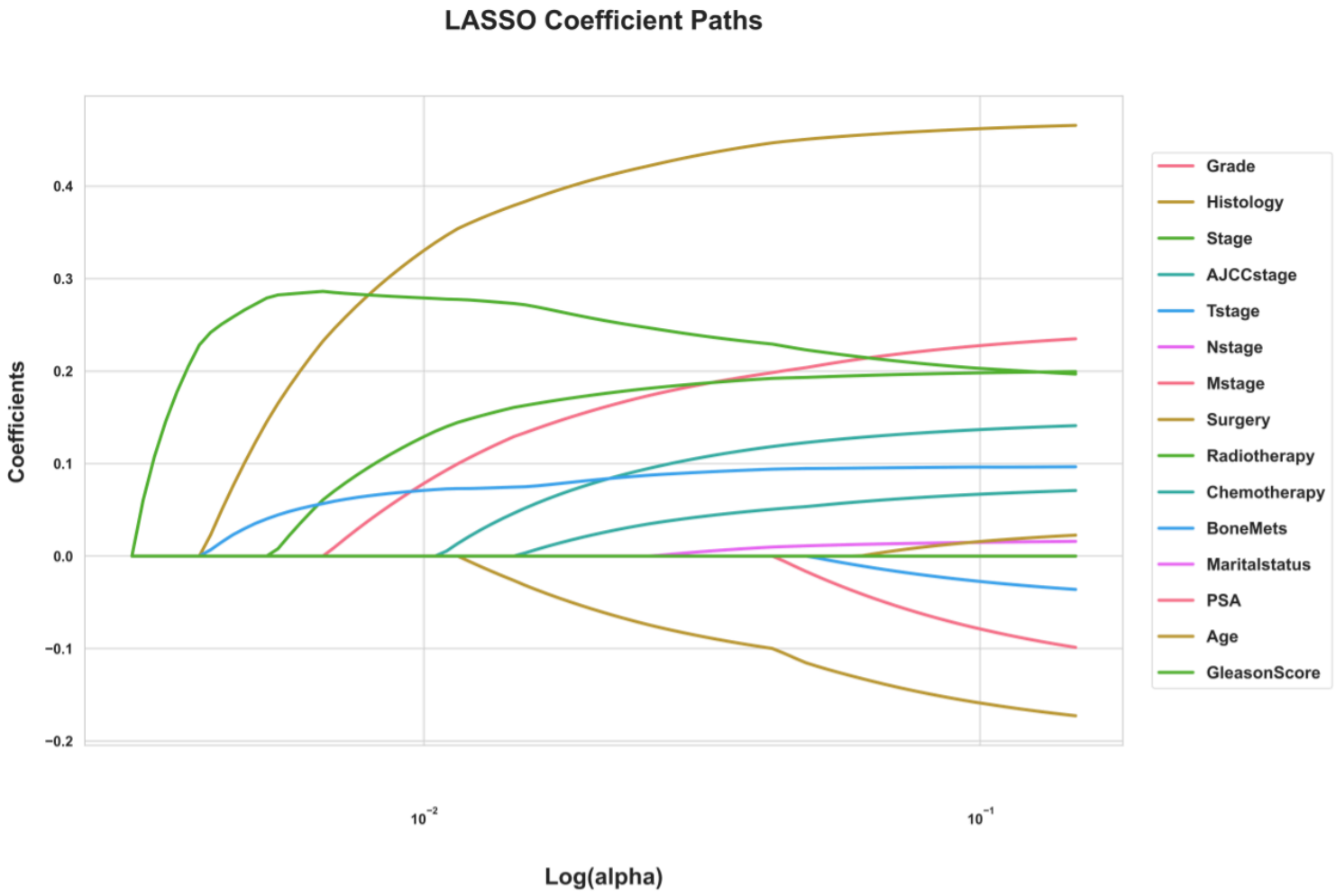


Figure 3

Coefficient paths of 15 clinical variables generated by LASSO regression for feature selection

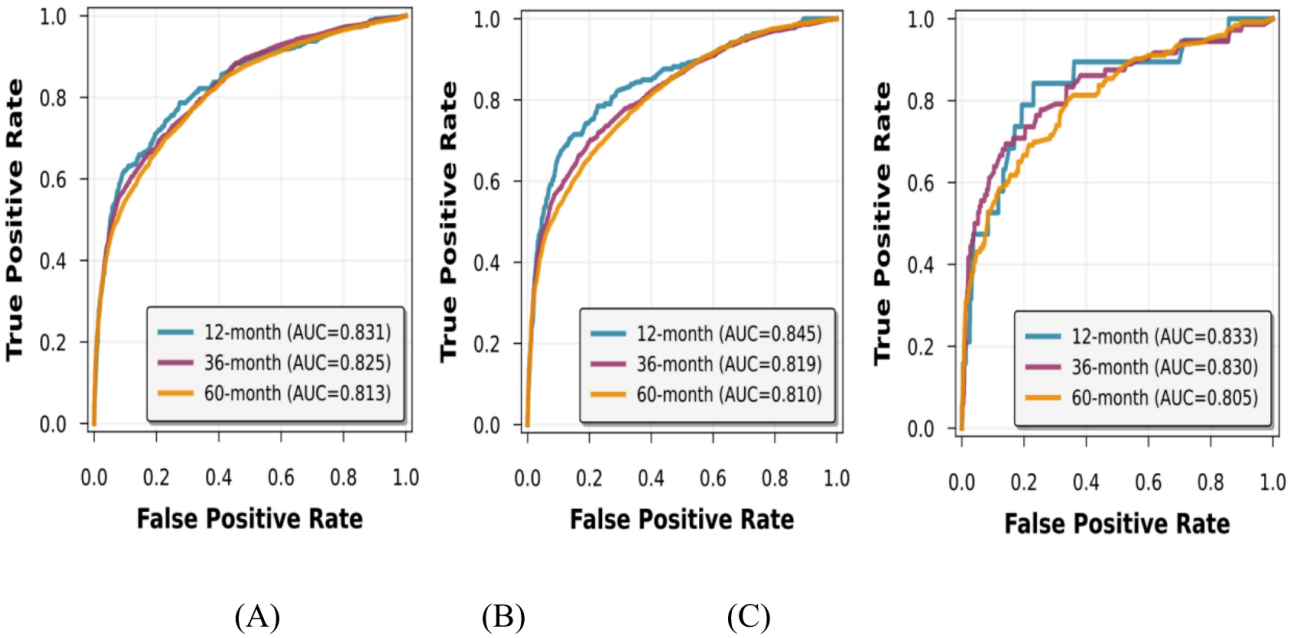


Figure 4

Time-dependent ROC curves of the boosted tree model for predicting 1-, 3-, and 5-year overall survival in the (A)training set, (B)internal validation set, and (C)external validation cohort

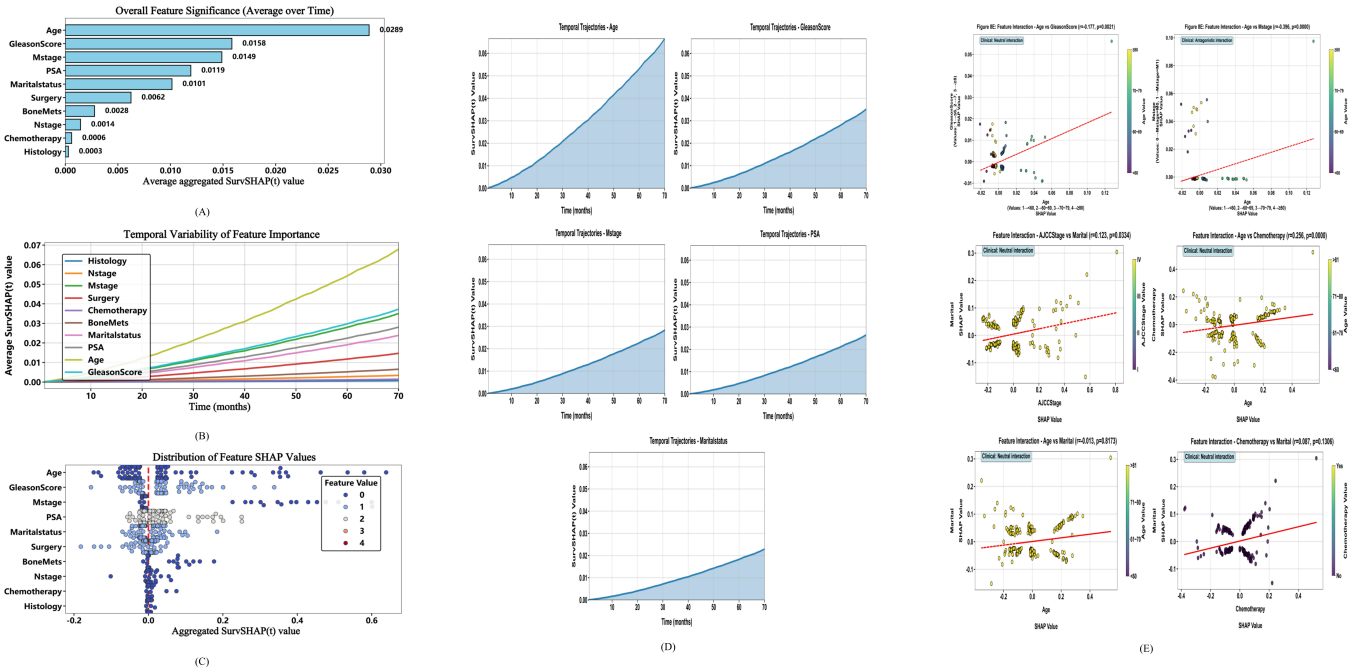
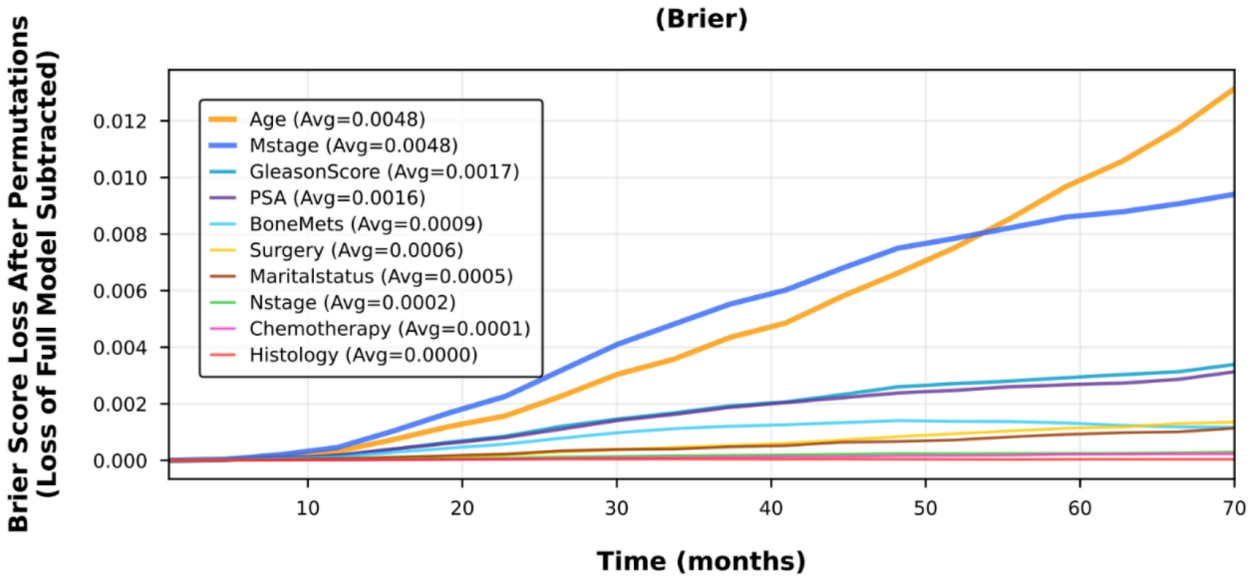
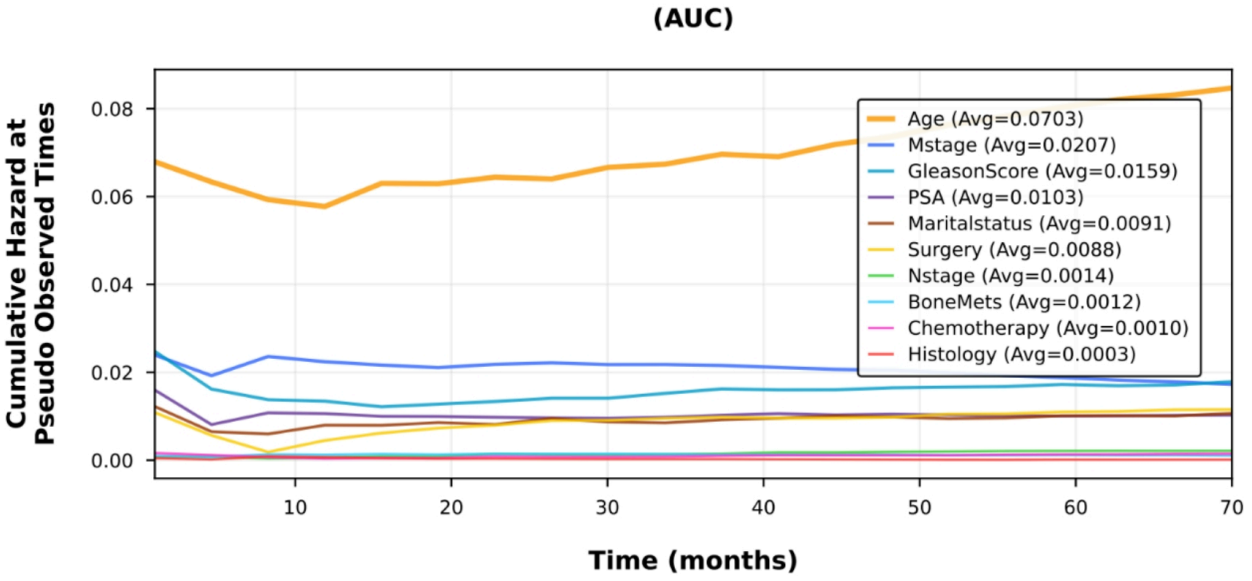


Figure 5

Time-varying permutation feature importance evaluated on the training cohort. (A) Increase in Brier score following feature permutation; (B) Decrease in C/D AUC following feature permutation. The y-axis denotes the magnitude of change in the respective loss metric after permuting each predictor



(A)



(B)

Figure 6

Time-varying permutation feature importance evaluated on the training cohort. (A) Increase in Brier score following feature permutation; (B) Decrease in C/D AUC following feature permutation. The y-axis denotes the magnitude of change in the respective loss metric after permuting each predictor

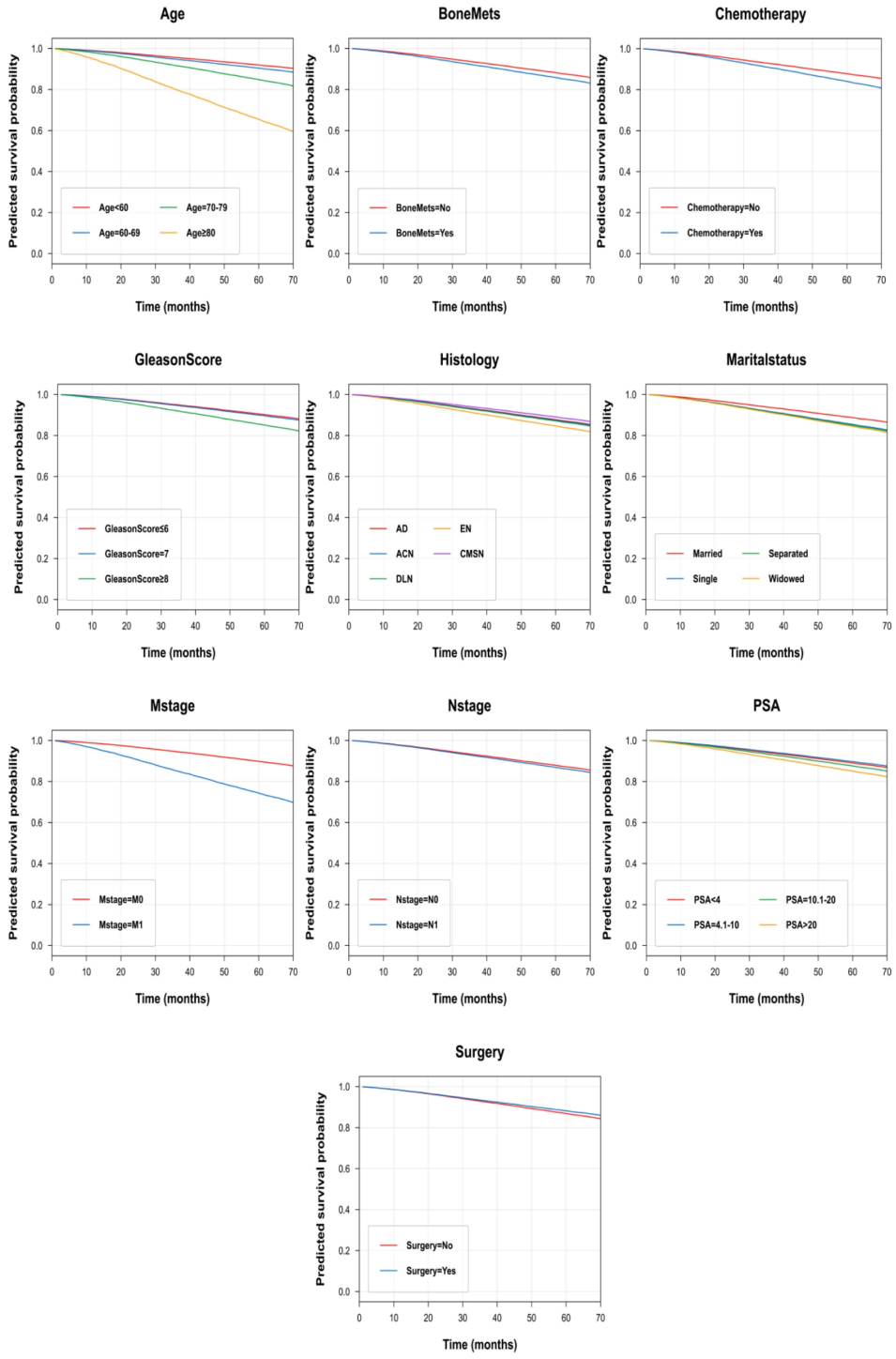


Figure 7

SurvSHAP(t) summary plots for global interpretation of feature contributions. (A) Bar chart depicting the overall importance ranking of all predictors based on mean absolute SurvSHAP(t) values. (B) Temporal variability of each variable's cumulative importance over follow-up time. (C) Bee swarm plot illustrating individual patient-level SurvSHAP(t) contributions: each point represents a specific feature value for a single patient, with the x-axis indicating the direction and magnitude of impact on model output, and features ordered by overall significance on the y-axis. (D) Time-dependent trajectory curves of the five most influential predictors. (E) Two-way interaction plots between selected feature pairs with correlation coefficients.

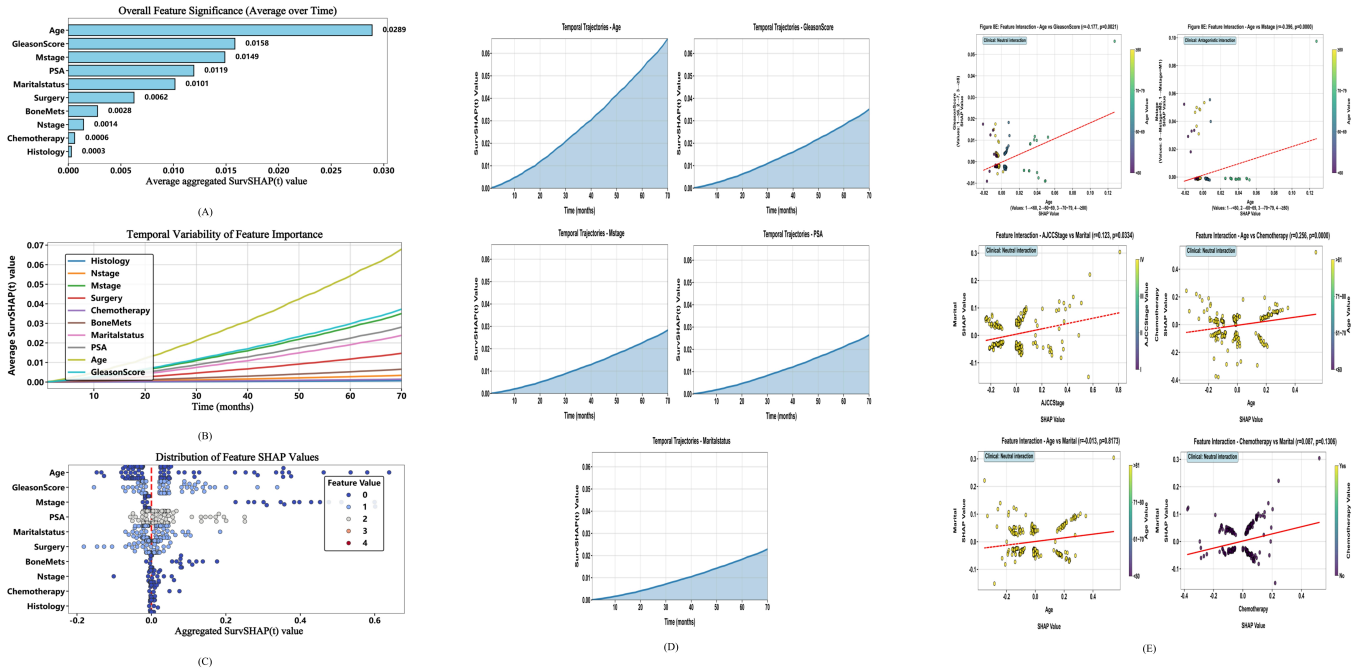


Figure 8

The SurvSHAP summary plot provides an overall interpretation of the global impact. (A) The length of the bar chart represents the overall significance of the variables. (B) The curve graph displays the cumulative importance of each variable. (C) Each point on the bee swarm plot represents a specific feature of a particular patient. The y-coordinate of the point is determined by the feature it represents, while the x-coordinate is determined by its impact on the model output. The features on the y-axis are sorted according to their significance. (D) Temporal trajectory curves of the top 5 influential features. (E) Two-way interaction plots of features.