

Supplementary Table

Supplementary Table 1 | Characteristics of the training and testing datasets

	Training	Prospective test	External test 1	External test 2
Sample size	25,899	4,131	747	170
CRT	1,031	80	44	36
Follow up time (days), Median (Q1, Q3)	9 (5-37)	5 (3-10)	3 (1-90)	57 (14.2-118)
General characteristics				
Age (years), Median (Q1, Q3)	59 (50-66)	60 (52-66)	56 (48-63)	59 (53.25-66)
Gender (male), <i>n</i> (%)	14202 (54.8%)	2272 (55%)	399 (53.4%)	37 (21.8%)
BMI (kg/m ²), Median (Q1, Q3)	23.38 (20.96-25.85)	23.46 (21.1-26.04)	22.77 (20.85-25.25)	24.91 (22.72-27.83)
Drink, <i>n</i> (%)	13015 (50.3%)	1495 (36.2%)	74 (9.9%)	37 (21.8%)
Smoke, <i>n</i> (%)	6864 (26.5%)	1688 (40.9%)	87 (11.6%)	37 (21.8%)
KPS, Median (Q1, Q3)	87.83 (80-90)	90 (80-90)	86.54 (78.97-88.29)	90 (80-90)
Comorbidity				
Diabetes, <i>n</i> (%)	3723 (14.4%)	658 (15.9%)	74 (9.9%)	42 (24.7%)
Hypertension, <i>n</i> (%)	8957 (34.6%)	1196 (29%)	142 (19%)	117 (68.8%)
Coronary.artery.disease, <i>n</i> (%)	1628 (6.3%)	306 (7.4%)	11 (1.5%)	3 (1.8%)
Cerebral.infarction, <i>n</i> (%)	829 (3.2%)	104 (2.5%)	6 (0.8%)	3 (1.8%)
Hyperlipidemia, <i>n</i> (%)	1459 (5.6%)	194 (4.7%)	276 (36.9%)	92 (54.1%)
Thrombosis.history, <i>n</i> (%)	1029 (4%)	188 (4.6%)	9 (1.2%)	4 (2.4%)
Blood markers				
ALB (g/L), Median (Q1, Q3)	38.21 (35.29-41.2)	38.7 (35.9-41.3)	39.3 (36.44-42.3)	45.1 (42.52-47.9)
Ddimer (mg/L), Median (Q1, Q3)	2.05 (0.61-5.73)	0.67 (0.34-1.46)	0.77 (0.47-1.66)	0.1 (0.06-0.16)
Hb (g/L), Median (Q1, Q3)	117.96 (105-129.83)	117 (106-128.5)	128 (115-139)	135 (127.25-143)
PLT (10 ⁹ /L), Median (Q1, Q3)	217 (162-281.18)	210 (160-269)	220.7 (179-278)	254 (216-300.75)
WBC (10 ⁹ /L), Median (Q1, Q3)	5.59 (3.84-7.89)	5.59 (4.26-)	5.31 (4.27-)	6.1 (4.78-)

		7.31)	6.79)	7.51)
Catheter characteristics				
CVC, <i>n</i> (%)	21296 (82.2%)	3699 (89.5%)	2 (0.3%)	61 (35.9%)
FICC, <i>n</i> (%)	794 (3.1%)	136 (3.3%)	0 (0%)	0 (0%)
PICC, <i>n</i> (%)	3597 (13.9%)	273 (6.6%)	498 (66.7%)	0 (0%)
PORT, <i>n</i> (%)	212 (0.8%)	23 (0.6%)	247 (33.1%)	109 (64.1%)
Improper.position, <i>n</i> (%)	401 (1.5%)	46 (1.1%)	36 (4.8%)	0 (0%)
Systematic treatment characteristics				
Nutrition, <i>n</i> (%)	4941 (19.1%)	568 (13.7%)	10 (1.3%)	61 (35.9%)
Immunotherapy, <i>n</i> (%)	2603 (10.1%)	757 (18.3%)	55 (7.4%)	0 (0%)
Anti.angiogenic.therapy, <i>n</i> (%)	2333 (9%)	435 (10.5%)	114 (15.3%)	0 (0%)
Anti.HER.mAb, <i>n</i> (%)	508 (2%)	138 (3.3%)	42 (5.6%)	93 (54.7%)
Anti.CD20.mAb, <i>n</i> (%)	478 (1.8%)	96 (2.3%)	0 (0%)	0 (0%)
Platinum, <i>n</i> (%)	14662 (56.6%)	1816 (44%)	627 (83.9%)	17 (10%)
Microtubule.inhibitor, <i>n</i> (%)	9424 (36.4%)	1095 (26.5%)	106 (14.2%)	107 (62.9%)
Antimetabolite, <i>n</i> (%)	7531 (29.1%)	954 (23.1%)	539 (72.2%)	0 (0%)
Antineoplastic.antibiotics, <i>n</i> (%)	4536 (17.5%)	765 (18.5%)	50 (6.7%)	156 (91.8%)
Nitrogen.mustard, <i>n</i> (%)	4472 (17.3%)	805 (19.5%)	52 (7%)	92 (54.1%)
DNA.topoisomerase.inhibitor, <i>n</i> (%)	3249 (12.5%)	488 (11.8%)	140 (18.7%)	0 (0%)
Cancer-related characteristics				
Stage.I, <i>n</i> (%)	838 (3.2%)	119 (2.9%)	33 (4.4%)	42 (24.7%)
Stage.II, <i>n</i> (%)	2849 (11%)	633 (15.3%)	113 (15.1%)	66 (38.8%)
Stage.III, <i>n</i> (%)	6067 (23.4%)	881 (21.3%)	342 (45.8%)	56 (32.9%)
Stage.IV, <i>n</i> (%)	16835 (65%)	2937 (71.1%)	259 (34.7%)	6 (3.5%)
Lung.pleura, <i>n</i> (%)	6760 (26.1%)	986 (23.9%)	113 (15.1%)	0 (0%)
Esophagus, <i>n</i> (%)	1878 (7.3%)	231 (5.6%)	19 (2.5%)	0 (0%)
Mediastinum, <i>n</i> (%)	144 (0.6%)	24 (0.6%)	0 (0%)	0 (0%)
Stomach, <i>n</i> (%)	2175 (8.4%)	347 (8.4%)	149 (19.9%)	17 (10%)
Intestine, <i>n</i> (%)	1957 (7.6%)	408 (9.9%)	366 (49%)	19 (11.2%)
Pancreas, <i>n</i> (%)	560 (2.2%)	97 (2.3%)	12 (1.6%)	0 (0%)
Liver, <i>n</i> (%)	394 (1.5%)	61 (1.5%)	4 (0.5%)	0 (0%)
BiliaryTract, <i>n</i> (%)	312 (1.2%)	32 (0.8%)	17 (2.3%)	0 (0%)
Renal.adrenal, <i>n</i> (%)	279 (1.1%)	46 (1.1%)	0 (0%)	0 (0%)
Peritoneum, <i>n</i> (%)	146 (0.6%)	31 (0.8%)	1 (0.1%)	0 (0%)
Rectum, <i>n</i> (%)	1913 (7.4%)	440 (10.7%)	118 (15.8%)	24 (14.1%)

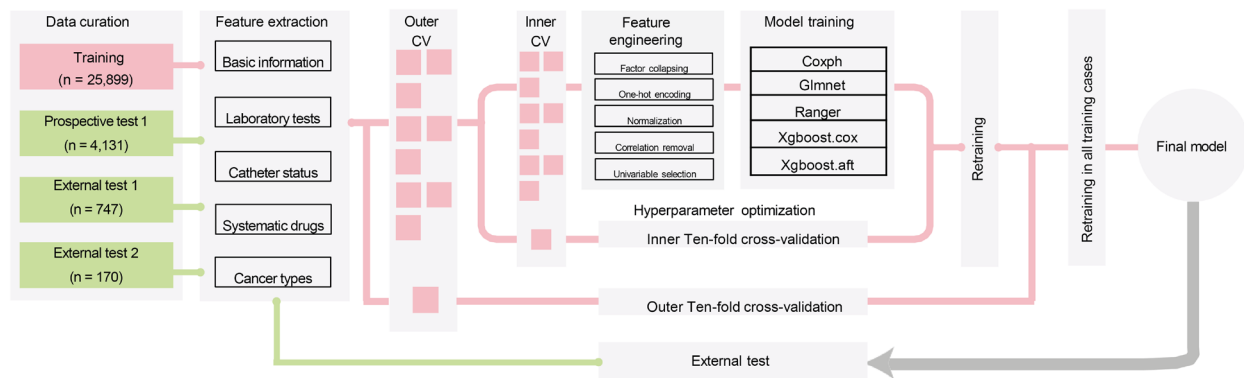
Bladder.ureter, <i>n</i> (%)	1351 (5.2%)	150 (3.6%)	2 (0.3%)	0 (0%)
Uterine.vagina, <i>n</i> (%)	819 (3.2%)	87 (2.1%)	5 (0.7%)	0 (0%)
Prostate, <i>n</i> (%)	540 (2.1%)	41 (1%)	0 (0%)	0 (0%)
Ovarian.fallopianTube, <i>n</i> (%)	504 (1.9%)	60 (1.5%)	8 (1.1%)	1 (0.6%)
Testis, <i>n</i> (%)	242 (0.9%)	31 (0.8%)	0 (0%)	0 (0%)
Breast, <i>n</i> (%)	4337 (16.7%)	765 (18.5%)	53 (7.1%)	109 (64.1%)
Head.neck, <i>n</i> (%)	317 (1.2%)	32 (0.8%)	0 (0%)	0 (0%)
TrunkWall, <i>n</i> (%)	224 (0.9%)	61 (1.5%)	0 (0%)	0 (0%)
Limb, <i>n</i> (%)	63 (0.2%)	8 (0.2%)	0 (0%)	0 (0%)
Lymphoma, <i>n</i> (%)	785 (3%)	154 (3.7%)	0 (0%)	0 (0%)

Supplementary Table 2 | Factor collapsing details for tumour types and chemotherapy types

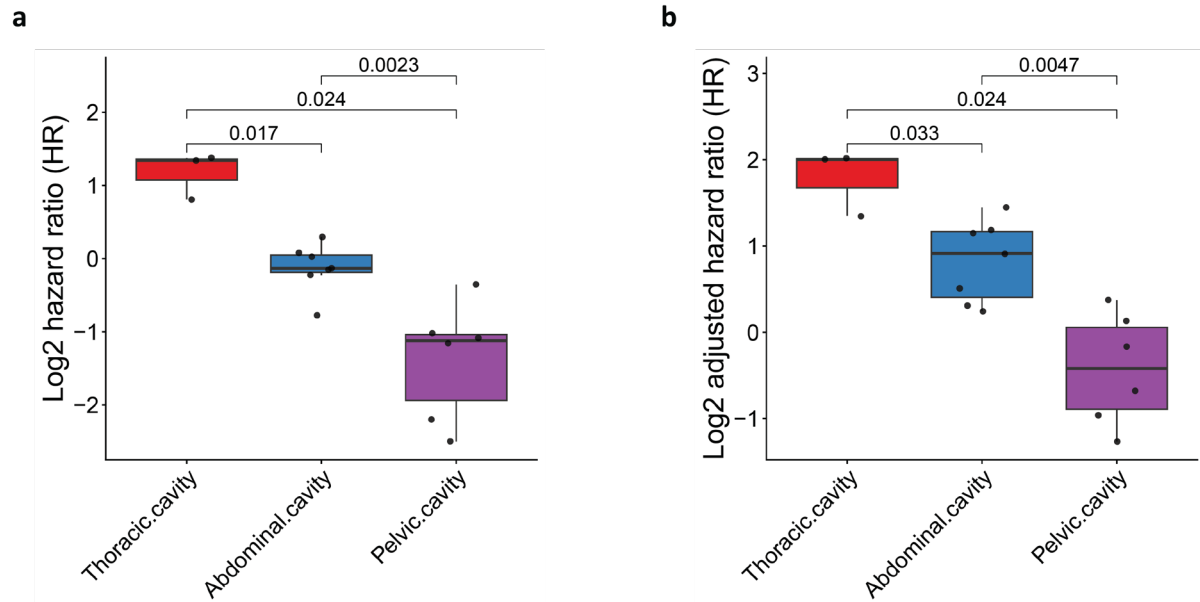
Variable types	Collapsed variables	Component variables
Tumour types	Lung.Pleura	Lung.cancer, Lung.Sarcoma, Lung.Neuroendocrine.tumor, Pleural.Sarcoma, Pleural.Neuroendocrine.tumor, Pleural.cancer
	Mediastinum	Mediastinal.Sarcoma, Mediastinal.Lymphoma, Mediastinal.Neuroendocrine.tumor, Mediastinal.cancer
	Esophagus	Esophageal.cancer, Esophageal.Neuroendocrine.tumor, Thymus.tumor
	Stomach	Stomach.Sarcoma, Stomach.Lymphoma, Stomach.Neuroendocrine.tumor, Stomach.cancer
	Intestine	Intestine.Sarcoma, Intestine.Lymphoma, Colorectal.Lymphoma, Colon.cancer, Duodenum.cancer, Intestine.tumor, Appendix.tumor
	Liver	Liver.Neuroendocrine.tumor, Liver.cancer
	BiliaryTract	Ampullary.cancer, Biliary.Tract.cancer
	Renal.Adrenal	Renal.cancer, Adrenal.tumor, Splenic.tumor
	Pancreas	Pancreatic.Neuroendocrine.tumor, Pancreatic.cancer
	Peritoneum	Peritoneal.Sarcoma, Peritoneal.Lymphoma, Peritoneal.Neuroendocrine.tumor, Peritoneal.tumor
	Rectum	Rectum.Sarcoma, Rectum.cancer
	Uterine.vagina	Uterine.Sarcoma, Cervical.Neuroendocrine.tumor, Cervical.cancer, Endometrial.cancer, Uterine.cancer, Vaginal.tumor
	Prostate	Prostate.Neuroendocrine.tumor, Prostate.cancer
	Ovarian.FallopianTube	Ovarian.cancer, Fallopian.tube.tumor, Pelvic.Sarcoma, Pelvic.tumor
	Testis	Testis.Lymphoma, Testis.cancer
	Bladder.ureter	Bladder.cancer, Urothelial.cancer
	Breast	Breast.cancer
	TrunkWall	Abdominal.wall.Sarcoma, Chest.wall.Sarcoma, Abdominal.wall.Skin.cancer, Vulvar.Skin.cancer, Back.Skin.cancer, Penile.Skin.cancer, Scrotal.Skin.cancer, Penis.cancer, Chest.wall.tumor, Back.tumor, Perianal.cancer, Abdominal.wall.tumor, Buttock.tumor,

		Spine.tumor	
Chemotherapy types	Limb	Upper.limb.Sarcoma, Upper.limb.Skin.cancer, Lower.limb.Sarcoma, Lower.limb.tumor	Upper.limb.Lymphoma, Upper.limb.tumor, Lower.limb.Skin.cancer,
	Head.neck	Head.and.neck.Sarcoma, Head.Neck.cancer	Head.and.neck.Lymphoma,
	Lymphoma	Lymphoma, Mediastinal.Lymphoma, Stomach.Lymphoma, Intestine.Lymphoma, Peritoneal.Lymphoma, Upper.limb.Lymphoma, Head.and.neck.Lymphoma	Colorectal.Lymphoma, Testis.Lymphoma,
	Antimetabolite	Pyrimidine.Analogues (including 5-Fluorouracil, Capecitabine, Cytarabine, Gemcitabine), Folate.Analogs (including Methotrexate, Pemetrexed, Pralatrexate), Purine.Analogs (including Fludarabine, Cladribine, Nelarabine)	
	Microtubule.Inhibitor	Taxanes, Epothilones, Vinca.Alkaloids, Eribulin	
	Nitrogen.mustard	Cyclophosphamide, Mechlorethamine, Melphalan, Ifosfamine, Bendamustine	
	Platinum	Cisplatin, Carboplatin, Oxaliplatin, Nedaplatin	
	DNA.topoisomerase.inhibitor	Epipodophyllotoxins, Camptothecin Analogs	
Antineoplastic.antibiotics		Anthracyclines, Bleomycin	

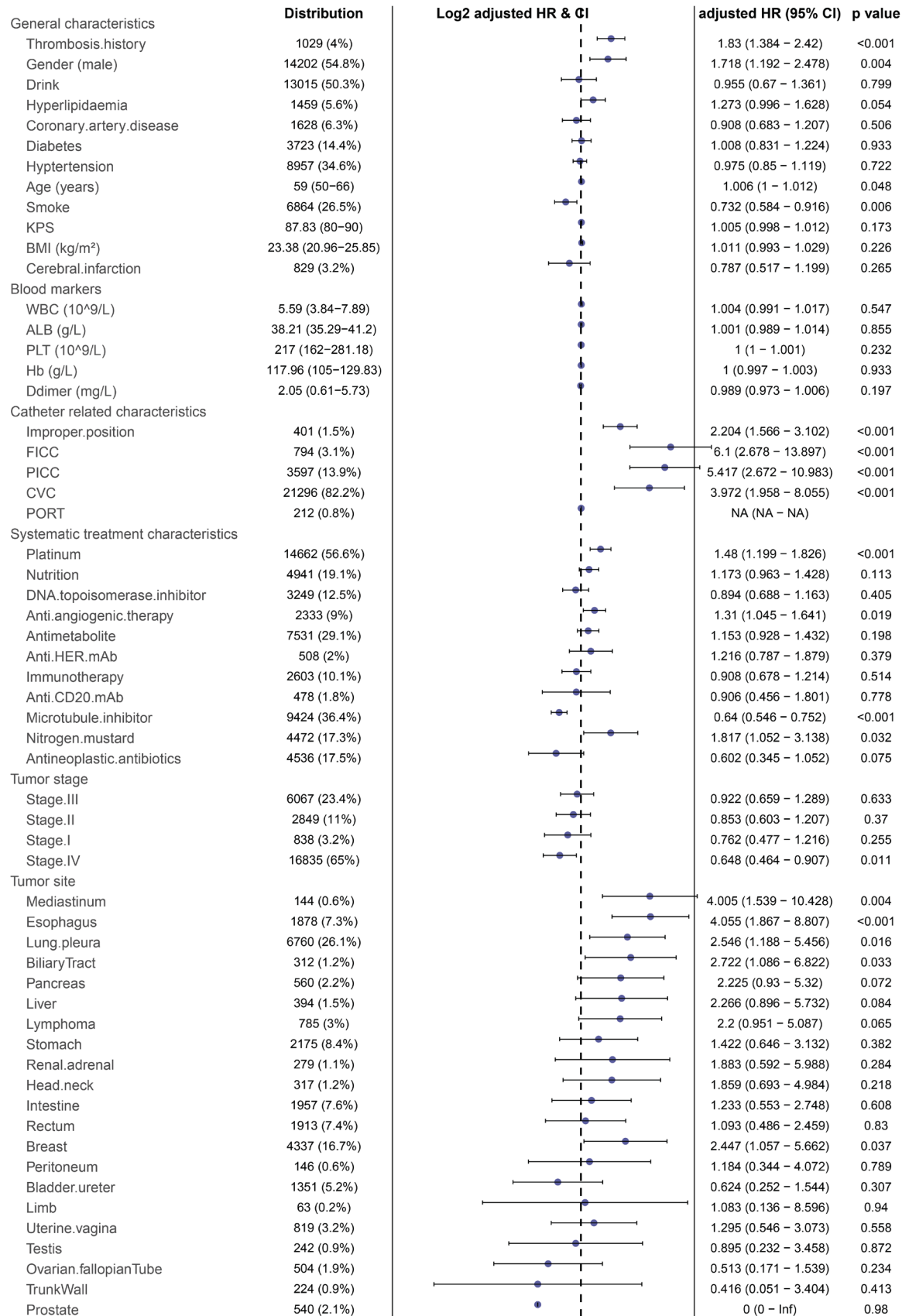
Supplementary Figures and Figure Legends



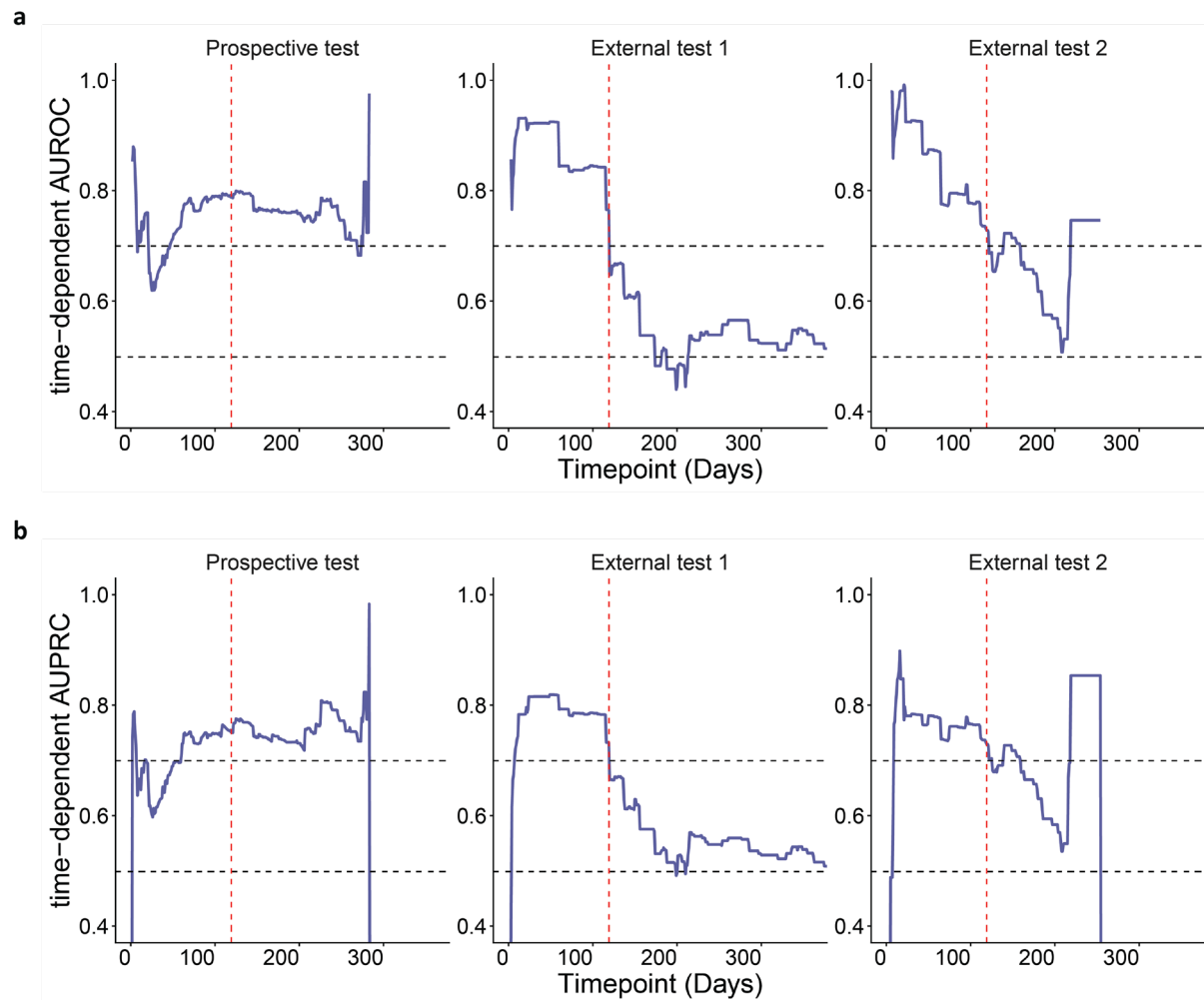
Supplementary Fig. 1 | Schematic of the machine learning survival model framework.
CV=cross-validation.



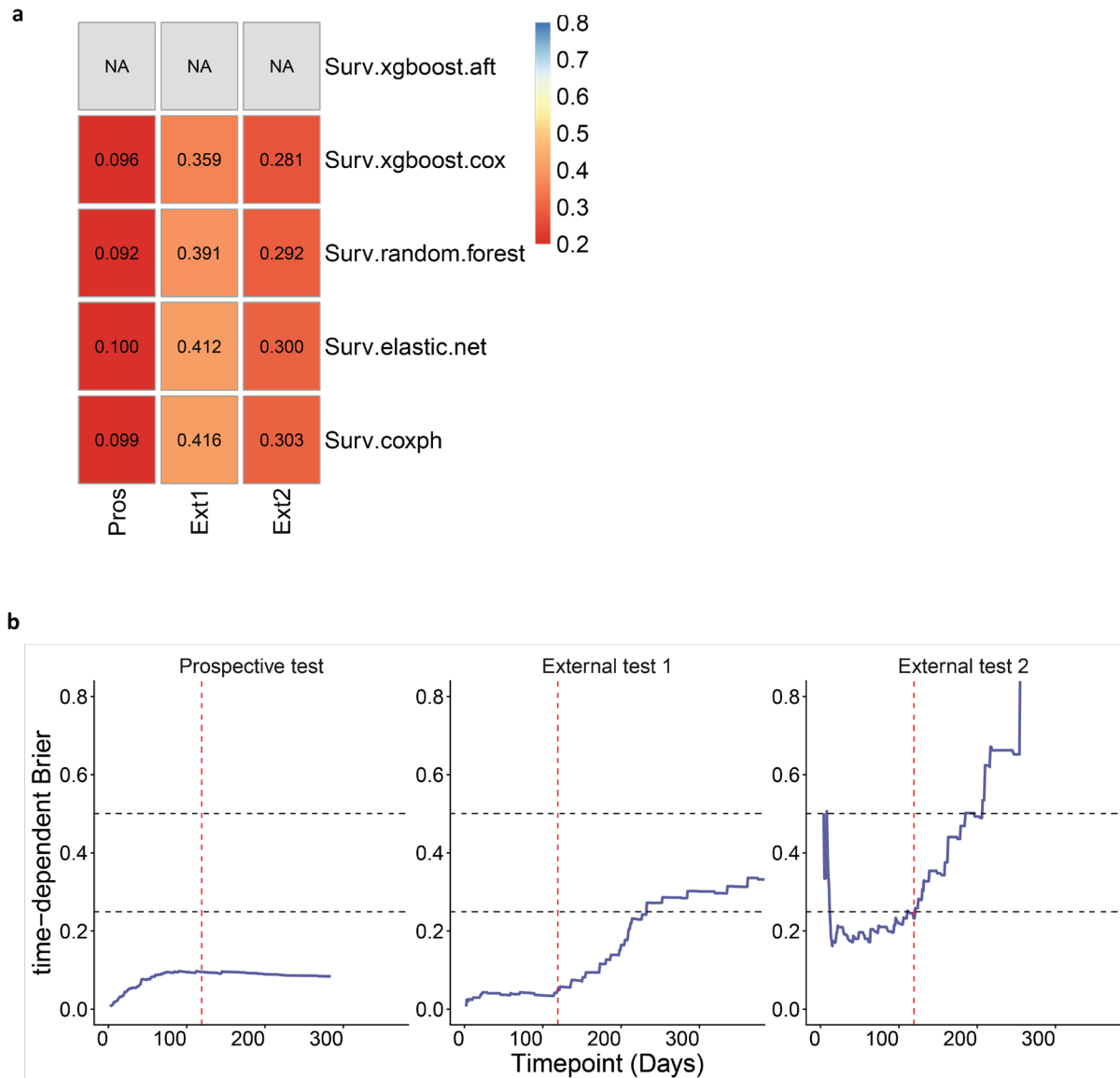
Supplementary Fig. 2 | Comparisons of the log2 hazard ratios between different tumour positions. a. Univariate Cox analysis hazard ratios. **b.** Multivariate Cox analysis adjusted hazard ratios. Log2 transformation is applied primarily for visualization purposes. The data are represented as boxplots, where the middle line is the median, the lower and upper hinges correspond to the 1st and 3rd quartiles, respectively, and the whisker corresponds to the minimum or maximum values no further than $1.5 \times \text{IQR}$ from the hinge (where the IQR is the interquartile range). Any data beyond the whiskers are considered outliers. P values were adjusted for multiple comparisons between cavities via the Benjamini–Hochberg (BH) procedure to control the false discovery rate (FDR). The positions are colour-coded.



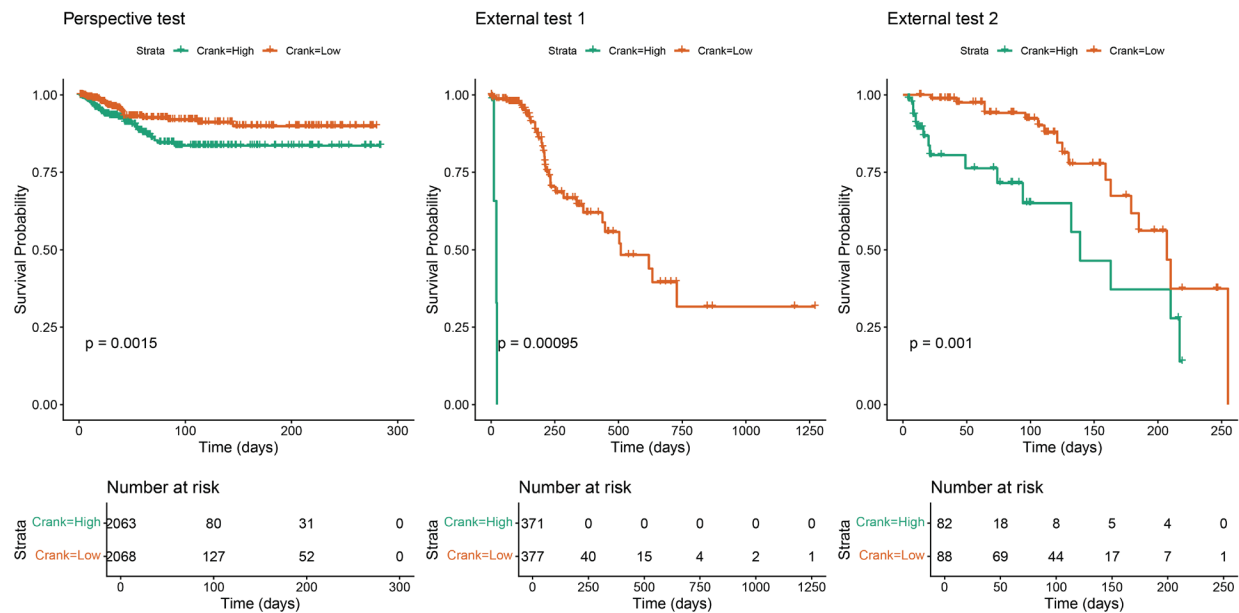
Supplementary Fig. 3 | Forest plot showing multivariate Cox analysis adjusted hazard ratios for CRT-free survival. HR=hazard ratio; CI=confidence interval; CRT=catheter-related thrombosis; BMI=body mass index; KPS=Karnofsky performance score; ALB=albumin; Hb=haemoglobin; PLT=platelets; WBC=white blood cells; CVC=central venous catheter; FICC=femorally inserted central catheter; PICC=peripherally inserted central catheter; PORT=implanted venous port; Anti.HER.mAb=human epidermal growth factor receptor monoclonal antibody; Anti.CD20.mAb=Anti-CD20 monoclonal antibody.



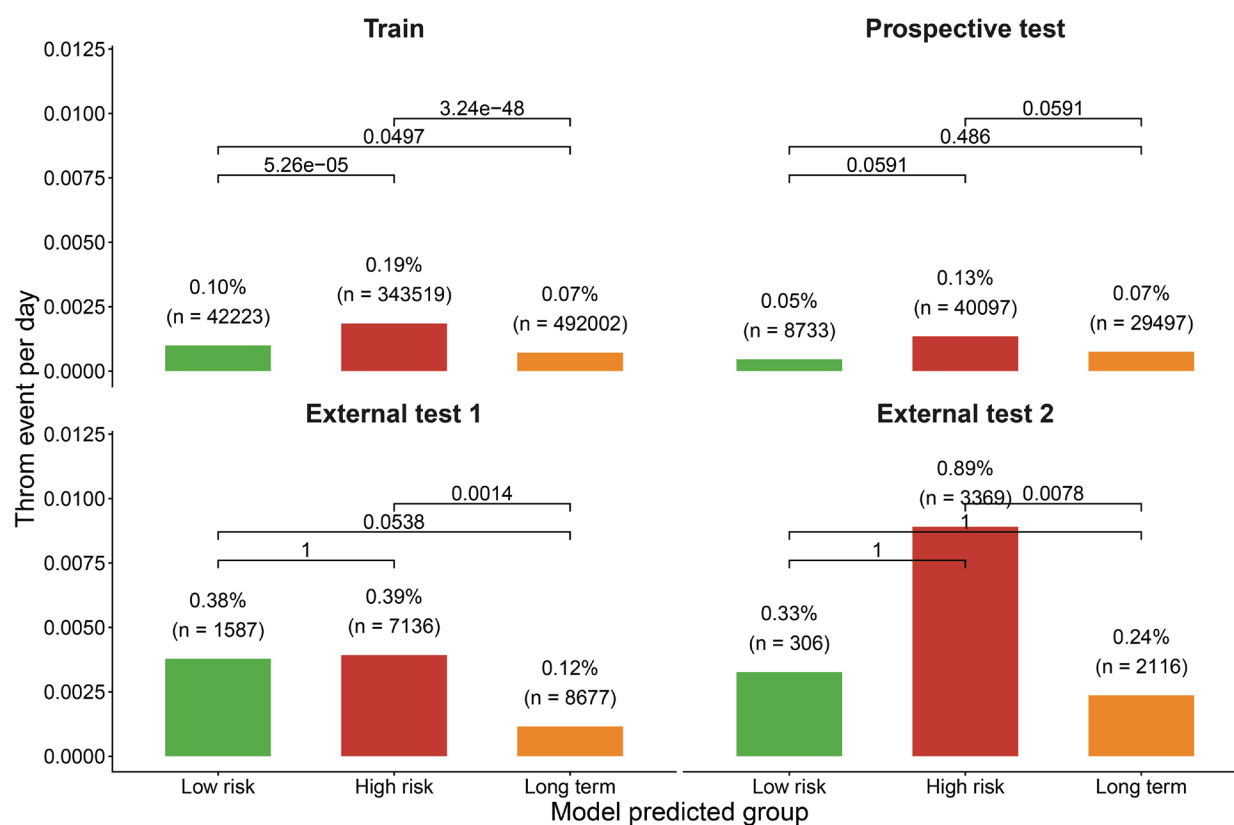
Supplementary Fig. 4 | Time-dependent computational performance of the trained survival models for catheter-related thrombosis (CRT). Time-dependent Area Under the Receiver Operating Characteristic Curve (AUROC) (**a**) and Area Under the Precision-Recall Curve (AUPRC) (**b**) of the SM-CRT model based on the three external test datasets. The 95th percentile of patient catheter insertion time (119 days) is marked as the vertical red dash line.



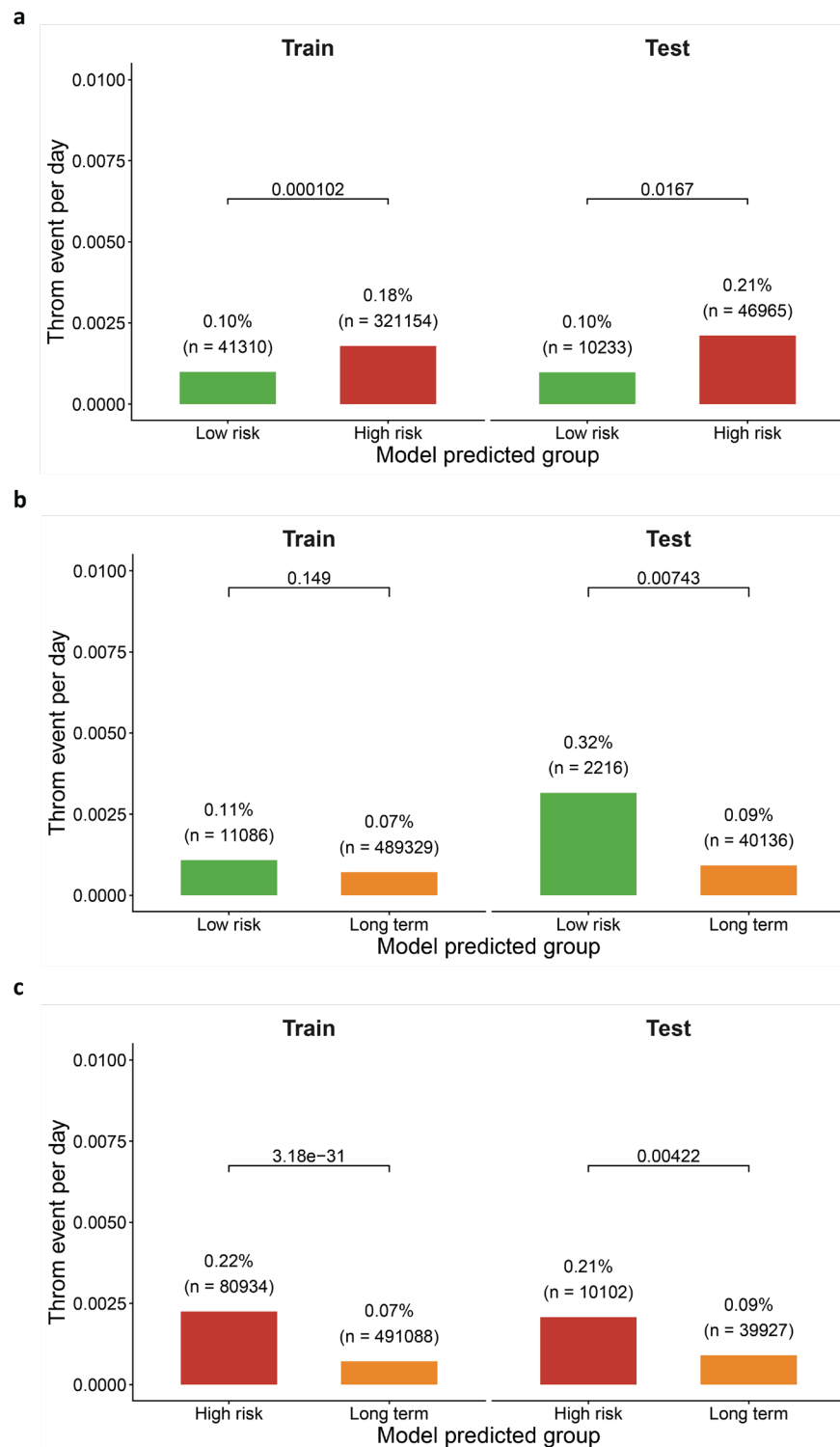
Supplementary Fig. 5 | Brier scores of the models based on the three external test datasets for calibration comparison of the developed survival models. a. Integrated Brier scores. The data are represented as a heatmap, where the integrated Brier score is colour coded and labelled in the grid of the heatmap. **b.** Time-dependent Brier scores. The 95th percentile of patient catheter insertion time (119 days) is marked as the vertical red dash line.



Supplementary Fig. 6 | Kaplan–Meier curves of CRT-free survival of patients based on the three external test datasets. Comparison between patients in the low crank group and high crank group split by the median crank value predicted by SM-CRT for each dataset. Crank value based stratifications are colour-coded.

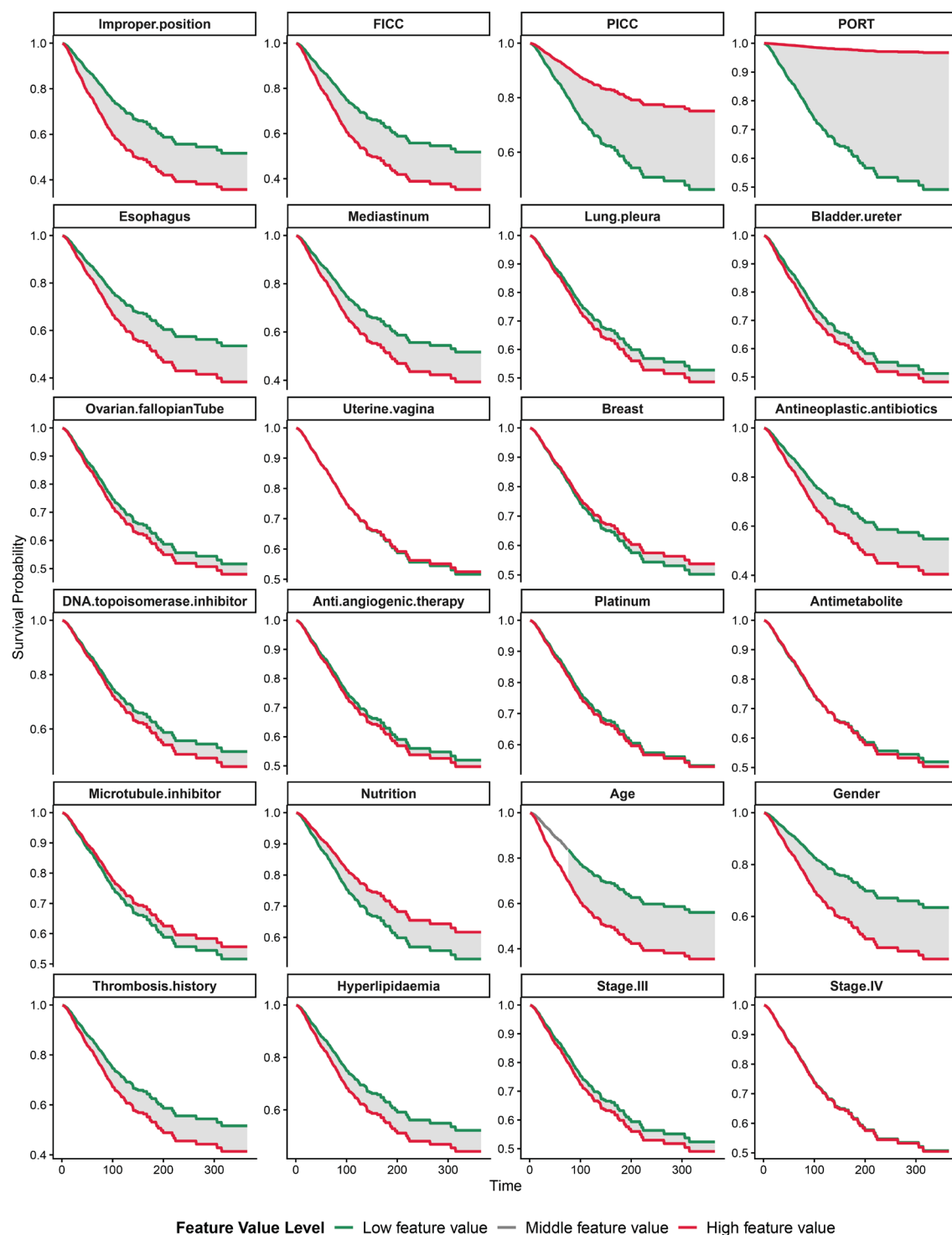


Supplementary Fig. 7 | Comparisons of the number of CRT events per day between different SM-CRT distr prediction-based groups via Fisher's exact test based on each dataset. The data are represented as histograms. P values were adjusted for multiple comparisons between groups via the Benjamini-Hochberg (BH) procedure to control the false discovery rate (FDR). The groups are colour-coded.



Supplementary Fig. 8 | Comparisons of the number of CRT events per day between different groups on the basis of the SM-CRT results via Fisher's exact test based on the training and

test datasets. Propensity score matching (PSM) was conducted to balance patient variables on the basis of the crank predictions by SM-CRT between pairs of low-risk and high-risk patients (**a**), low-risk and long-term patients (**b**), and high-risk and long-term patients (**c**). The data are represented as histograms. The groups are colour-coded.



Supplementary Fig. 9 | Dataset-level variable profile as a partial dependence explanation (partial dependence plot [PDP]) for survival models to explore the model response as a function of the selected variables. Multiple Kaplan–Meier curves are used to display response

changes as the variable value changes. The area between the highest and lowest time–survival function curves was calculated to determine the importance of each feature.