doi 10.12122/j.issn.1673-4254.2023.06.10 J South Med Univ, 2023, 43(6): 952-963

机器学习模型和Cox回归模型预测食管胃结合部腺癌预后的效能

高凯绩，王一豪，曹海坤，贾建光
蚌埠医学院第一附属医院肿瘤外科，安徽 蚌埠 233000

摘要：目的 探讨机器学习和传统Cox回归模型在预测食管胃结合部腺癌（AEG）患者术后生存能力中的应用价值。**方法** 选取2015年9月~2020年10月本院收治的287例AEG患者，排除失访及临床资料缺失者，共筛选出203例患者的临床病理资料，经过对数据的赋值等处理，转换成满足R语言分析数据的要求的数据。将203例患者数据使用随机数表法按照3∶1的比例划分为训练集和验证集，对两组数据分别进行Cox比例风险模型构建和4种机器学习模型的构建，绘制出ROC曲线、校准曲线和临床决策曲线（DCA）。为评估4种机器学习模型之间的预测效能，进行机器学习模型的内部验证。通过曲线下面积(AUC)评价模型预测的性能，校准曲线反映模型的拟合情况，并通过DAC判断其临床意义。**结果** Cox等比例风险回归、极端梯度提升、随机森林、支持向量机、多层感知机验证集中3年生存率的AUC值分别为0.870、0.901、0.791、0.832、0.725，验证集中5年生存率的AUC值分别为0.915、0.916、0.758、0.905、0.737。4种机器学习模型内部验证分别是：极端梯度提升（AUC=0.818）、随机森林（AUC=0.772）、支持向量机（AUC=0.804）、多层感知机（AUC=0.745）。**结论** 机器学习模型对于AEG患者生存率预测的表现优于Cox等比例风险回归模型，尤其在不满足等比例假设或线性回归模型下，并能够包含较多的影响变量。在内部验证中，XGBoost模型的预测效能最好，支持向量机次之，随机森林出现过拟合，多层感知机受数据量影响可能拟合效果较差。
关键词：食管胃结合腺癌；人工智能；机器学习；Cox比例风险回归模型

Efficacy of machine learning models *versus* Cox regression model for predicting prognosis of esophagogastric junction adenocarcinoma

GAO Kaiji, WANG Yihao, CAO Haikun, JIA Jianguang
Department of Surgical Oncology, First Affiliated Hospital of Bengbu Medical College, Bengbu 233000, China

Abstract: Objective To compare the performance of machine learning models and traditional Cox regression model in predicting postoperative outcomes of patients with esophagogastric junction adenocarcinoma (AEG). **Methods** This study was conducted among 203 AEG patients with complete clinical and follow-up data, who were treated in our hospital between September, 2015 and October, 2020. The clinicopathological data of the patients were processed for analysis using R language package and divided into training and validation datasets at the ratio of 3∶1. The Cox proportional hazards regression model and 4 machine learning models were constructed for analyzing the datasets. ROC curves, calibration curves and clinical decision curves (DCA) were plotted. Internal validation of the machine learning models was performed to assess their predictive efficacy. The predictive performance of each model was evaluated by calculating the area under the curve (AUC), and the model fitting was assessed using the calibration curve. **Results** For predicting 3-year survival based on the validation dataset, the AUC was 0.870 for Cox proportional hazard regression model, 0.901 for eXtreme Gradient Boosting (XGBoost), 0.791 for random forest, 0.832 for support vector machine, and 0.725 for multilayer perceptron; For predicting 5-year survival, the AUCs of these models were 0.915, 0.916, 0.758, 0.905, and 0.737, respectively. For internal validation, the AUCs of the 4 machine learning models decreased in the order of XGBoost (0.818), random forest (0.758), support vector machine (0.0.804), and multilayer perceptron (0.745). **Conclusion** The machine learning models show better predictive efficacy for survival outcomes of patients with AEG than Cox proportional hazard regression model, especially when proportional odds assumption or linear regression models are not applicable. XGBoost models have better performance than the other machine learning models, and the multi-layer perception model may have poor fitting results for a limited data volume.
Keywords: esophagogastric junction adenocarcinoma; artificial intelligence; machine learning; Cox proportional hazard regression model

目前，各国报道的食管胃结合部腺癌(AEG)发病率均呈一定上升趋势[1-3]。中国、日本及其他亚洲国家亦有类似的研究结果[4]。因此，这类肿瘤引起了更多学者的关注和重视。AEG具有胃癌和食管癌的基本特性，但又有所不同，其淋巴结转移即可上至胸腔纵膈又可下至腹腔，肿瘤位置处于食管胃交界处，手术难度大，操作复杂。临床外科对于该病手术治疗的预后认知不足，且AEG患者在临床病理分期、治疗方案等方面存在不同，其预后差异很大。

Cox比例风险回归模型（Cox-PH）通常用于队列研究[5]，以确定风险因素，并使用生存数据构建预测模型。

收稿日期：2022-11-12
基金项目：蚌埠医学院512人才培育计划（by51202207）
作者简介：高凯绩，在读硕士研究生，E-mail: 1252609038@qq.com
通信作者：贾建光，教授，硕士生导师，E-mail: Jiajianguang1978@126.com

Cox-PH可以根据风险比值来定义变量的重要性，具有直观、解释性强的特点。然而，Cox-PH受到等比例风险和线性等假设的限制，若假设不满足则影响模型预测效果。此外，Cox-PH对于非线性复杂关系的变量难以处理[6]。机器学习（ML）是让计算机通过学习处理过的数据和信息，开发出一种算法，通过算法让机器学习如何做出决策。ML为通过经验自动改进的计算机算法[7]，包括XGBoost、随机森林（RF）、支持向量机（SVM）、多层感知机（MLP）等，它们有能力从数据中学习训练以达到对未来事件的准确预测[8]，这些算法正逐渐被用于肺癌、乳腺癌、肝癌、胃肠道癌等多种恶性肿瘤预后研究，成为临床研究的热点[9-12]。ML模型包含种类较多，且每种模型都具有自己的适用范围，不同种类、不同体量、不同特征的数据其预测效能也有所不同，传统的Cox-PH和ML模型对AEG预后预测效能孰优孰劣尚未了解。因此，本文旨在构建Cox-PH和ML模型，比较两类模型对AEG患者生存的预测效能。

1 资料和方法

1.1 资料收集

选取2015年9月~2020年10月蚌埠医学院第一附属医院收治的287例AEG患者。根据世界卫生组织将AEG定义为：肿瘤中心位于食管胃结合部（EGJ）上下5 cm范围内，且肿瘤本身必须跨越或者直接接触EGJ，病理为腺癌的肿瘤[13]。排除标准：不符合世界卫生组织对AEG定义的临床数据；临床资料不全者；合并有严重心肺功能疾病者；术后失访者；未经手术治疗；死亡原因不明者。剔除后，共筛选出203例。采用电话随访的方法进行了为期5年的随访，时间截止2022年4月。

1.2 数据处理

查阅相关文献，找出可能影响AEG预后的相关因素，包括患者术前性别、年龄、Borrmann分型、分化程度、浸润深度(T分期) 、淋巴结转移数量(N分期)、病理TNM分期、肿瘤最大直径、术后化疗、纤维蛋白原（Fibr）、D二聚体（D-dimer）、手术方式、术后住院时间、营养指数（PNI）、中性粒细胞计数与淋巴细胞计数的比值(NLR)、白球比（WBR）、血清癌胚抗原(CEA)、甲胎蛋白（AFP）、糖类抗原199（CA199）等，共19项观察指标。分析前因年龄、肿瘤大小、Fibr、D-dimer、PNI、术后住院时间、NLR、WRB、CEA、AFP、CA199等11个连续性变量数值过多，容易造成模型过拟合，所以本研究采用X-tile3.6.1软件（耶鲁大学团队研发）[14]，根据所需参数要求，对这11个连续变量取最佳截断值。将以上19项可能影响AEG预后的观察指标量化处理，转化成满足SPSS 26.0、R4.2.2、Rstudio2022和python3.11.0处理的数据进行分析。

1.3 训练集和验证集的划分

采用生存分析方法Kaplan-Meier法进行生存曲线绘制（图1），判断变量是否符合比例风险假设，并估计不同影响因素下AEG的生存率，使用Log-rank检验组间差异是否有统计学意义（表1）。在变量满足比例风险假设检验条件下，对变量进行单因素Cox回归分析，并对单因素分析中具有统计学意义（P<0.05）的变量进行分析，将其作为协变量纳入多因素Cox回归中，得到多因素分析中具有统计学意义（P<0.05）的变量（表2），将203例患者数据依据单因素Cox回归分析的变量使用随机数表法按照3:1的比例分为训练集153例和验证集50例，卡方检验用于比较训练集和验证集之间的分类变量，连续性变量采用均数±标准差进行统计描述分析（表3）。

1.4 模型构建

1.4.1 Cox比例风险模型构建 将训练集中Cox多因素分析有统计学意义的变量作为独立预后因素纳入构建Cox-PH。通过逐步回归法计算风险比率（HR）和95%可信区间（CI），并以列线图的形式显展示。根据列线图（图2），计算训练集和验证集中各项得分总和得到总分，将总分按照列线图对应生存率，绘制出训练集和验证集3年和5年的ROC曲线、校正曲线以及临床决策曲线（DCA）（图3、图4和图7）。

1.4.2 ML模型构建 153例训练集数据用于建模，本研究基于Cox单因素分析中的13个变量（P<0.05）作为输入。在训练集中利用网格搜索方法确定XGBoost、随机森林（RF）、支持向量机（VM）、多层感知机（MLP）模型最优的超参数，逻辑回归采用默认参数。并使用5折重采样验证对模型进行训练和内部验证，最终得到最佳参数。各模型参数如下，XGBoost模型：优化目标函数使用二元logistic，学习速率设置为0.3，最大树深度设置为8；最小分叉权重和设置为4，L2正则化系数设置为1。RF模型：度量指标使用gini，最小分叉纯度收益设置为0.0，树数目设置为100。SVM模型：正则化因子设置为1.0，核类型使用rbf，收敛度量设置为0.1。MLP模型：输入层、隐藏层1、隐藏层2、输出层的节点分别设置为13、20、10、1，每一层运算方式都采用relu激活函数并随机初始化。训练次数为20次。经过多次训练计算出训练集中各模型3年和5年的AUC值，并在验证集中予以验证。为评估4种ML模型之间的预测效能，使用该4种ML模型对所有样本进行生存预测，在按照7:3比例随机划分独立训练集和验证集后，对训练集应用交叉验证进行超参数调优，充分利用训练集数据，避免测试集数据的泄露，同时在验证集中验证，统计各个模型在训练集和测试集上AUC表现。

1.5 统计学方法

本研究连续变量使用X-til3.6.1进行最佳截断值选

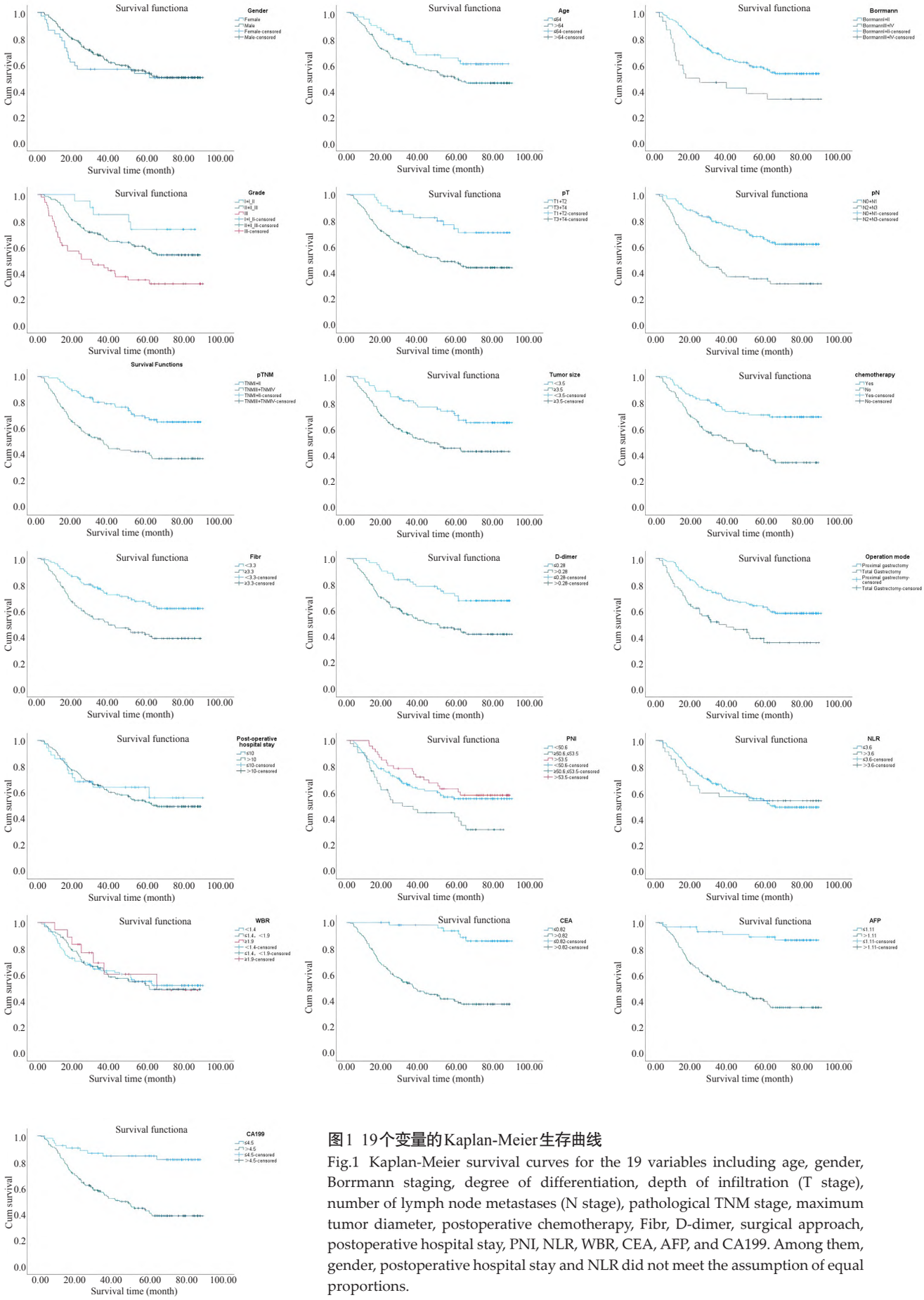

图1 19个变量的Kaplan-Meier生存曲线

Fig.1 Kaplan-Meier survival curves for the 19 variables including age, gender, Borrmann staging, degree of differentiation, depth of infiltration (T stage), number of lymph node metastases (N stage), pathological TNM stage, maximum tumor diameter, postoperative chemotherapy, Fibr, D-dimer, surgical approach, postoperative hospital stay, PNI, NLR, WBR, CEA, AFP, and CA199. Among them, gender, postoperative hospital stay and NLR did not meet the assumption of equal proportions.

表1 Log-rank检验

Tab.1 Log-rank test of the clinicopathological data of the patients

Variable	*n*	Survival time (months, *Mean*±*SD*)	χ^2	*P*
Gender			0.414	0.520
Female	37	52.855±6.040		
Male	166	58.261±2.649		
Age (years)			4.048	0.044
≤64	60	64.034±4.073		
>64	143	54.188±2.946		
Borrmann			8.567	0.003
I+II	173	59.462±2.517		
III+IV	30	41.859±6.705		
Grade			18.281	<0.001
I+I-II	20	71.723±5.106		
II+II-III	134	59.988±2.836		
III	49	41.370±5.078		
pT			10.694	0.001
T1+T2	47	72.334±3.849		
T3+T4	156	52.633±2.839		
pN			26.475	<0.001
N0+N1	125	66.414±2.713		
N2+N3	78	41.919±3.947		
pTNM			19.626	<0.001
I+II	94	68.581±2.983		
III+IV	109	47.004±3.414		
Tumor size (cm)			11.284	0.001
<3.5	71	69.329±3.482		
≥3.5	132	50.410±3.064		
Chemotherapy			19.133	<0.001
Yes	90	68.930±3.32		
No	113	47.436±3.170		
Fibr (g/L)			13.163	<0.001
<3.3	102	66.364±3.131		
≥3.3	101	47.963±3.473		
D-dimer (mg/L)			12.600	<0.001
≤0.28	62	70.271±3.445		
>0.28	141	51.170±3.028		
Surgical approach			12.778	<0.001
Proximal gastrectomy	129	63.639±2.837		
Total gastrectomy	74	45.554±4.194		
Postoperative hospital stay (d)			0.110	0.741
≤10	36	59.918±6.313		
>10	167	56.544±2.596		
PNI			8.233	0.016
<50.6	112	59.526±3.335		
≥50.6,≤53.5	43	44.156±4.966		
>53.5	47	64.169±4.326		
NLR			0.001	0.975
≤3.6	168	56.984±2.598		
>3.6	35	55.507±6.257		
WBR			0.155	0.925
<1.4	88	57.330±3.780		
≥1.4, <1.9	97	56.135±3.408		
≥1.9	18	58.876±7.839		
CEA (ng/mL)			34.647	<0.001
≤0.82	54	83.681±1.993		
>0.82	149	47.518±2.817		
AFP (ng/mL)			34.871	0.005
≤1.11	58	79.718±2.777		
>1.11	145	47.807±2.855		
CA199 (U/mL)			21.836	<0.001
≤4.5	54	75.832±3.561		
>4.5	149	50.319±2.827		

表2 Cox单因素、多因素分析

Tab.2 Cox univariate and multifactor analysis of the clinicopathological data of the patients

Factor	Univariate analysis HR (95% *CI*)	*P*	Mulivariate analysis HR (95% *CI*)	*P*
Gender		0.520		
Female	1.00			
Male	0.85 (0.51-1.41)			
Age (year)		0.046		0.879
≤64	1.00		1.00	
>64	1.64 (1.01-2.67)		1.04 (0.61-1.79)	
Borrmann		0.004		0.009
I+II	1.00		1.00	
III+IV	2.09 (1.26-3.46)		2.16 (1.21-3.84)	
Grade		<0.001		0.216
I+I-II	1.00		1.00	
II+II-III	2.05 (0.82-5.11)	0.125	1.36 (0.52-3.55)	0.528
III	4.45 (1.73-11.43)	0.002	1.98 (0.73-5.40)	0.183
pT		0.002		
T1+T2	1.00			
T3+T4	2.65 (1.44-4.86)			
pN		<0.001		
N0+N1	1.00			
N2+N3	2.81 (1.86-4.23)			
pTNM		<0.001		0.004
I+II	1.00		1.00	
III+IV	2.61 (1.68-4.04)		1.99 (1.25-3.18)	
Tumor size (cm)		0.001		0.021
<3.5	1.00		1.00	
≥3.5	2.20 (1.37-3.54)		1.85 (1.10-3.13)	
Chemotherapy		<0.001		<0.001
Yes	1.00		1.00	
No	2.66 (1.69-4.19)		3.01 (1.79-5.07)	
Fibr (g/L)		<0.001		0.978
<3.3	1.00		1.00	
≥3.3	2.14 (1.40-3.26)		1.01 (0.63-1.62)	
D-dimer (mg/L)		0.001		0.124
≤0.28	1.00		1.00	
>0.28	2.43 (1.47-4.03)		1.55 (0.89-2.70)	
Surgical approach		<0.001		0.043
Proximal gastrectomy	1.00		1.00	
Total gastrectomy	2.09 (1.38-3.16)		1.61 (1.02-2.54)	
Postoperative hospital stay (d)		0.741		
≤10	1.00			
>10	1.10 (0.61-1.99)			
PNI		0.019		0.069
<50.6	1.00		1.00	
≥50.6, ≤53.5	1.76 (1.09-2.83)	0.020	1.88 (1.10-3.20)	
>53.5	0.82 (0.48-1.40)	0.458	1.30 (0.71-2.38)	
NLR		0.975		
≤3.6	1.00			
>3.6	1.01 (0.59-1.73)			
WBR		0.870		
<1.4	1.00			
≥1.4, <1.9	1.52 (0.85-2.153)			
≥1.9	0.97 (0.70-1.35)			
CEA (ng/mL)		<0.001		0.003
≤0.82	1.00		1.00	
>0.82	8.13(3.55-18.63)		3.58 (1.52-8.42)	
AFP (ng/mL)		<0.001		0.004
≤1.11	1.00		1.00	
>1.11	7.36 (3.39-15.95)		3.33 (1.46-7.59)	
CA199 (U/mL)		<0.001		0.201
≤4.5	1.00		1.00	
>4.5	4.48 (2.25-8.93)		1.61 (0.78-3.33)	

表3 Cox单因素分析结果(*P*<0.05)分为训练集和验证集

Tab.3 Division of the results of Cox univariate analysis (*P*<0.05) into training and validation datasets [*n* (%)]

Variables	Training set (*n*=153)	Validation set (*n*=50)	*P*
Survival (month, *Mean*±*SD*)	45.742±2.209	43.7992±3.876	2.496
Age (year)			
≤64	46 (30.1%)	14 (28%)	0.921
>64	107 (69.9%)	36 (72%)	
Borrmann			
I+II	128 (83.7%)	45 (90%)	0.386
III+IV	25 (16.3%)	5 (10%)	
pT			
T1+T2	33 (21.6%)	14 (28%)	0.458
T3+T4	120 (78.4%)	36 (72%)	
pN			
N0+N1	95 (62.1%)	30 (60%)	0.923
N2+N3	58 (37.9%)	20 (40%)	
pTNM			
I+II	71 (46.4%)	23 (46%)	1.000
III+IV	82 (53.6%)	27 (54%)	
Tumor size			
<3.5 cm	52 (34%)	19 (38%)	0.730
≥3.5 cm	101 (66%)	31 (62%)	
Chemotherapy			
yes	69 (45.1%)	21 (42%)	0.827
No	84 (54.9%)	29 (58%)	
Fibr (g/L)			
<3.3	80 (52.3%)	22 (44%)	0.393
≥3.3	73 (47.7%)	28 (56%)	
D-dimer (mg/L)			
≤0.28	49 (32%)	13 (26%)	0.531
>0.28	104 (68%)	37 (74%)	
Surgical approach			
Proximal gastrectomy	99 (64.7%)	30 (60%)	0.666
Total gastrectomy	54 (35.3%)	20 (40%)	
CEA (ng/mL)			
≤0.82	38 (24.8%)	16 (32%)	0.417
>0.82	115 (75.2%)	34 (68%)	
AFP (ng/mL)			
≤1.11	47 (30.7%)	11 (22%)	0.315
>1.11	106 (69.3%)	39 (78%)	
CA199 (U/mL)			
≤4.5	40 (26.1%)	14 (28%)	0.941
>4.5	113 (73.9%)	36 (72%)	
Grade			
I+I-II	15 (9.8%)	5 (10%)	1.000
II+II-III	101 (66%)	33 (66%)	
III	37 (24.2%)	12 (24%)	
PNI			
<50.6	89 (58.2%)	24 (48%)	0.444
≥50.6, ≤53.5	31 (20.3%)	12 (24%)	
>53.5	33 (21.6%)	14 (28%)	

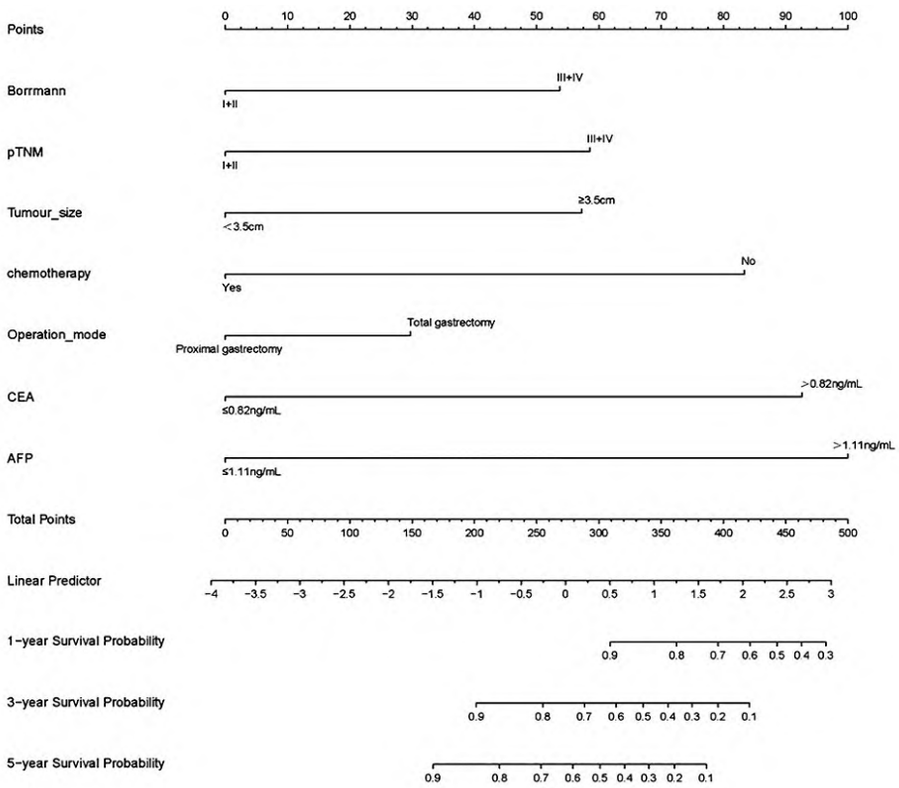

图2 训练集中Cox多因素分析有统计学意义的变量作为独立预后因素纳入构建Cox-PH绘制出的列线图

Fig.2 Nomogram constructed based on the Cox-PH from training set variables included as independent prognostic factors that were significant in Cox multifactor analysis (Borrmann staging, pathological TNM stage, maximum tumor diameter, postoperative chemotherapy, surgical approach, CEA, AFP). The risk ratio (hazard ratio, HR) and 95% confidence interval (CI) were calculated by stepwise regression method and displayed as a columnar line graph (Nomogram).

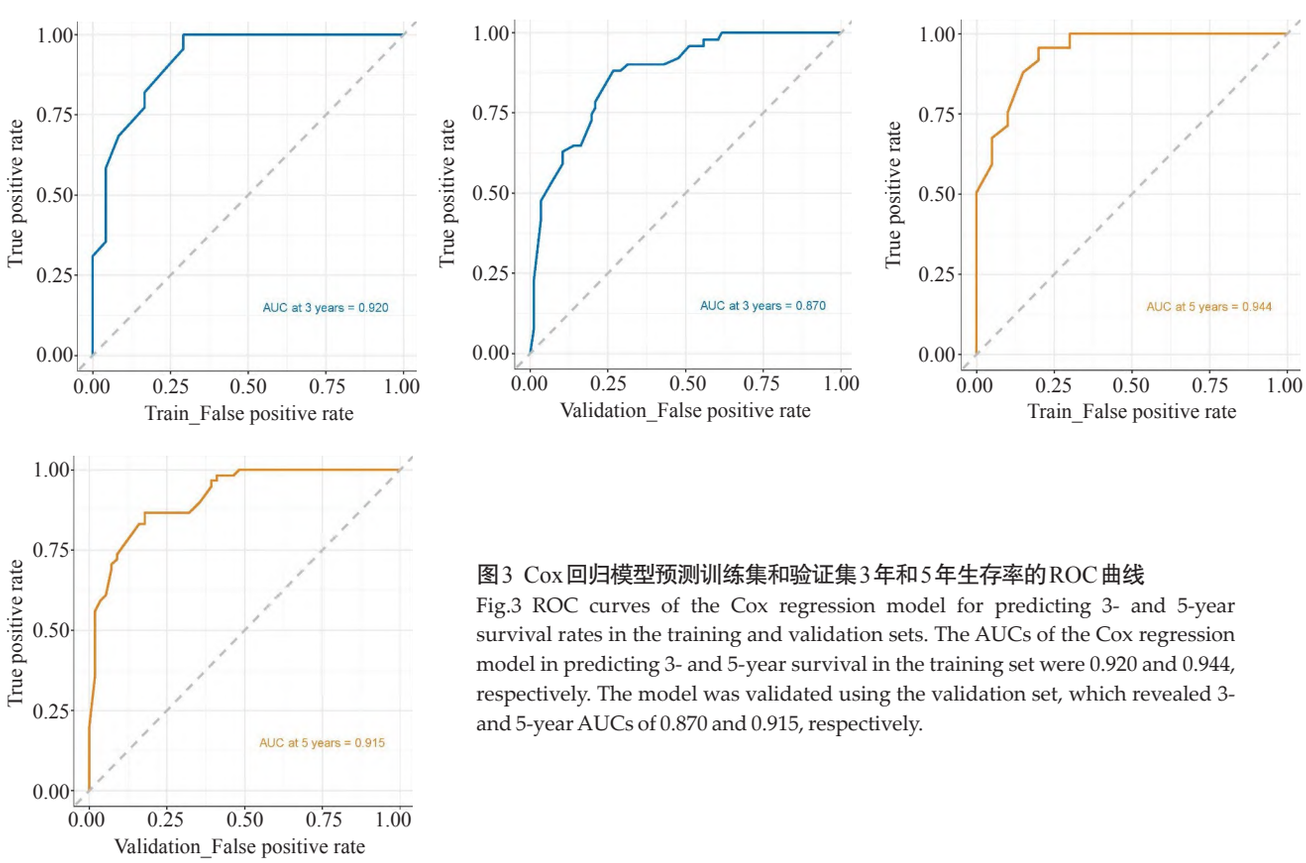

图3 Cox回归模型预测训练集和验证集3年和5年生存率的ROC曲线

Fig.3 ROC curves of the Cox regression model for predicting 3- and 5-year survival rates in the training and validation sets. The AUCs of the Cox regression model in predicting 3- and 5-year survival in the training set were 0.920 and 0.944, respectively. The model was validated using the validation set, which revealed 3- and 5-year AUCs of 0.870 and 0.915, respectively.

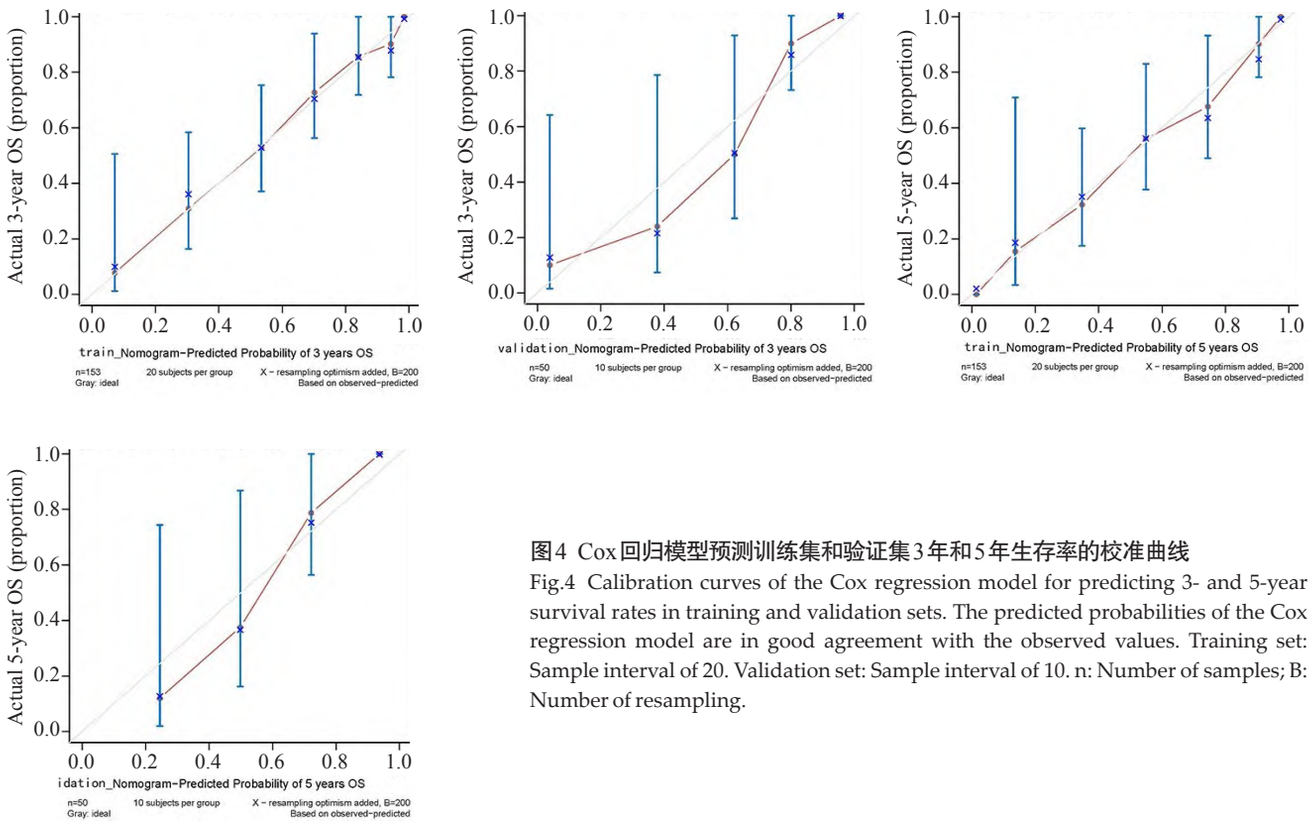

图4 Cox回归模型预测训练集和验证集3年和5年生存率的校准曲线

Fig.4 Calibration curves of the Cox regression model for predicting 3- and 5-year survival rates in training and validation sets. The predicted probabilities of the Cox regression model are in good agreement with the observed values. Training set: Sample interval of 20. Validation set: Sample interval of 10. n: Number of samples; B: Number of resampling.

取。数据分析使用SPSS 26.0软件，计数资料使用t检验法，计量资料使用卡方或者精确概率分析法，Kaplan-Meier法统计各分组生存率，log-rank法对比各分组之间的差异，P<0.05为差异具有统计学意义。Cox回归模型、机器学习模型构建、图形绘制、程序开发采用R软件4.2.2版本（R Foundation for Statistical Computing, Vienna, Austria, http:/www.Rproject.org/）、Rstudio（2022, PBC, Boston, MA, http:/www.rstudio.com/）和python3.11.0。

2 结果

2.1 患者一般资料

本研究共纳入AEG患者203例，其中男性166例，女性37例，男女比为4.48∶1，平均年龄为67.9岁，年龄范围为44~83岁，≥60岁占82.84%。11个连续变量通过X-tile进行最佳截断值选取，分别为：年龄（64岁）、肿瘤大小（3.5 cm）、Fibr（3.3 g/L）、D-dimer（0.28 mg/L）、PNI（50.6, 53.5）、术后住院时间（10 d）、NLR（3.6）、WRB（1.4, 1.9）、CEA（0.82 ng/mL）、AFP（1.11 ng/mL）、CA199（4.5 U/mL），分组结果见表1。

2.2 Cox-PH预测效能

本研究采用Kaplan-Meier法对19个变量进行生存曲线绘制（图1）。根据生存曲线，其中性别、术后住院时间和NLR不符合等比例假设，Log-rank检验显示其中年龄（P=0.044）、Borrmann分型（P=0.003）、分化程度（P<0.001）、浸润深度（P=0.001）、淋巴结转移数量（P<0.001）、病理TNM分期（P<0.001）、肿瘤最大直径（P=0.001）、术后化疗（P<0.001）、Fibr（P<0.001）、D-dimer（P<0.001）、手术方式（P<0.001）、PNI（P=0.016）、CEA（P<0.001）、AFP（P=0.005）、CA199（P<0.001）共15个变量具有统计学意义（表1）。基于Cox单因素分析结果P<0.05的15个变量，排除浸润深度和淋巴转移数，进行Cox多因素分析，获得7个有统计学意义的变量分别为：Borrmann分型（P=0.009）、病理TNM分期（P=0.004）、肿瘤最大直径（P=0.021）、术后化疗（P<0.001）、手术方式（P=0.043）、CEA（P=0.003）、AFP（P=0.004，表2）。Cox回归模型预测训练集3年和5年生存率的AUC分别为0.920、0.944。使用验证集予以验证：验证集3年AUC=0.870、验证集5年AUC=0.915。Cox回归模型训练集和验证集的临床决策曲线（DCA）（图7）显示：在模型的阈值设置在10%~90%阈值范围内，决策曲线位于None线和All线的上方，因此该模型具有临床实用性。校准曲线提示，Cox回归模型的预测概率与实际观察值具有较好的一致性。

2.3 机器学习模型预测效能

训练集中机器学习模型对3年预测的AUC值分别为：XGBoost（AUC=0.913）、RF（AUC=0.997）、SVM（AUC=0.954）、MLP（AUC=0.701）；5年预测的AUC值

分别为：XGBoost（AUC=0.922）、RF（AUC=0.999）、SVM（AUC=0.953）、MLP（AUC=0.784）。验证集对模型予以验证，3年的预测AUC值分别为：XGBoost（AUC=0.901）、RF（AUC=0.791）、SVM（AUC=0.832）、MLP（AUC=0.725）；5年的预测AUC值分别为：XGBoost（AUC=0.916）、RF（AUC=0.758）、SVM（AUC=0.905）、MLP（AUC=0.737）（表4）。4种ML模型基于全部样本独立划分训练集，预测所有时间段生存率结果如下：XGBoost（AUC=0.900）、RF（AUC=0.999）、SVM（AUC=0.928）、MLP（AUC=0.781）；验证集对模型予以验证，生存率结果如下：XGBoost（AUC=0.818）、RF（AUC=0.772）、SVM（AUC=0.804）、MLP（AUC=0.745）（表5）。通过绘制4种模型验证集预测的临床决策曲线（图8），4种ML模型验证集ROC分析中模型的最佳截断值分别为XGBoost（cutoff=38.5%）、RF（cutoff=50.7%）、MLP（cutoff=46.1%）、SVM（cutoff=46.0%），在该阈值下，XGBoost和SVM模型的决策曲线位于None线和All线的上方，MLP和RF模型的决策曲线没有或不完全位于None线和All线的上方，故XGBoost和SVM模型具有临床实用性。综合4种ML模型验证集预测的AUC得分森林图、校准曲线（图5、6）和临床决策曲线。可以看出XGBoost模型稳定性较好；SVM预测效能其次；RF模型的预测概率与实际观察值的一致性较差，可能出现过拟合，即训练集中变现出色，但在验证集中变现不佳；MLPL模型的预测概率和实际观察值均不理想。

表4 各机器学习模型在训练集及验证集中3年和5年预测表现的对比

Tab.4 Comparison of predictive performance of the 4 machine learning models for 3-year and 5-year survival of the patients in the training and validation sets

Model	Training set				Validation set			
	AUC	AC	SE	SP	AUC	AC	SE	SP
XGB								
3 years	0.913	0.810	0.852	0.795	0.901	0.772	0.900	0.710
5 years	0.922	0.845	0.848	0.855	0.916	0.872	0.957	0.853
RF								
3 years	0.997	0.965	1.000	0.944	0.791	0.672	0.667	0.800
5 years	0.999	0.979	1.000	0.963	0.758	0.623	1.000	0.455
SVM								
3 years	0.954	0.873	0.944	0.841	0.832	0.801	0.780	0.610
5 years	0.953	0.894	0.984	0.838	0.905	0.821	0.926	0.588
MLP								
3 years	0.701	0.606	0.255	0.813	0.725	0.541	0.348	0.684
5 years	0.784	0.711	0.763	0.699	0.737	0.656	0.900	0.548

XGB: eXtreme gradient boosting; RF: Random forests; SVM: Support vector machines; MLP: Multi-layer perceptron; AC: Accuracy; SE: Sensitivity; SP: Specificity.

表5 各ML模型在训练集及验证集的预测表现

Tab.5 Prediction performance of each machine learning model in the training and validation sets

Model	Training Set					Validation Set				
	AUC	Cutoff	AC	SE	SP	AUC	Cutoff	AC	SE	SP
XGB	0.9	0.453	0.812	0.883	0.76	0.818	0.385	0.727	0.85	0.764
RF	0.999	0.503	0.977	1	0.959	0.772	0.507	0.689	0.791	0.755
MLP	0.781	0.459	0.721	0.705	0.743	0.745	0.461	0.658	0.818	0.709
SVM	0.928	0.464	0.85	0.865	0.848	0.804	0.46	0.733	0.828	0.755

3 讨论

AEG早期确诊率低，绝大部分就诊时已有淋巴结转移[15,16]，且由于其解剖部位特殊，与传统的远端胃癌的比较，AEG的总体预后较差。根据文献报道，手术根治率平均在80%左右，根治术后的5年生存率也仅为30%左右[17,18]。本研究资料显示AEG术后3年生存率为43.8%，5年生存率为37.0%，其5年生存率略低于他人报道的贲门癌5年生存率40%[19]，近似于研究报道的AEG5年生存率38.5%[20]，因此本研究为预测AEG术后生存率提供了新的参考。

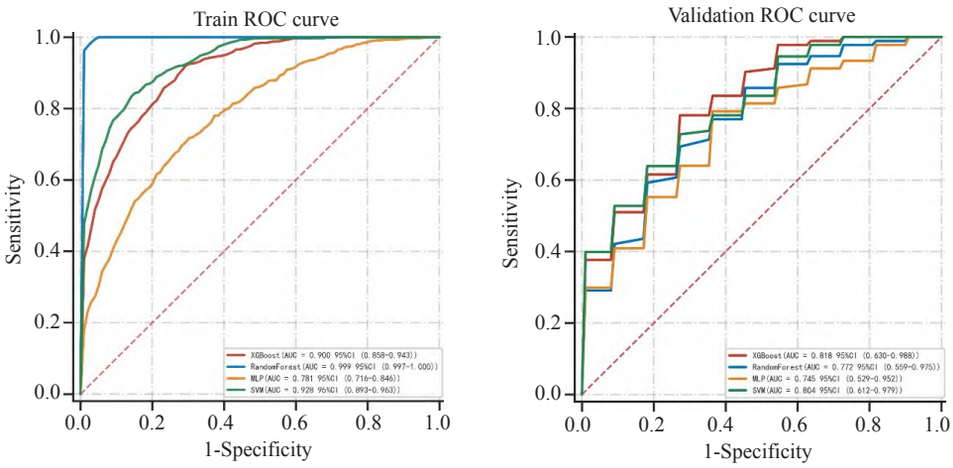

图5 4种ML模型预测训练集和验证集生存率的ROC曲线

Fig.5 Survival prediction ROC curves of the 4 machine learning models for the training and validation sets. In the training set, the predictive performance of survival outcomes for each independent model was: XGBoost (AUC=0.900), RF (AUC=0.999), SVM (AUC=0.928), and MLP (AUC=0.781) while for the validation set, the predictive performance was: XGBoost (AUC=0.818), RF (AUC=0.772), SVM (AUC=0.804), MLP (AUC=0.781). RF (AUC=0.772), SVM (AUC=0.804), and MLP (AUC=0.745).

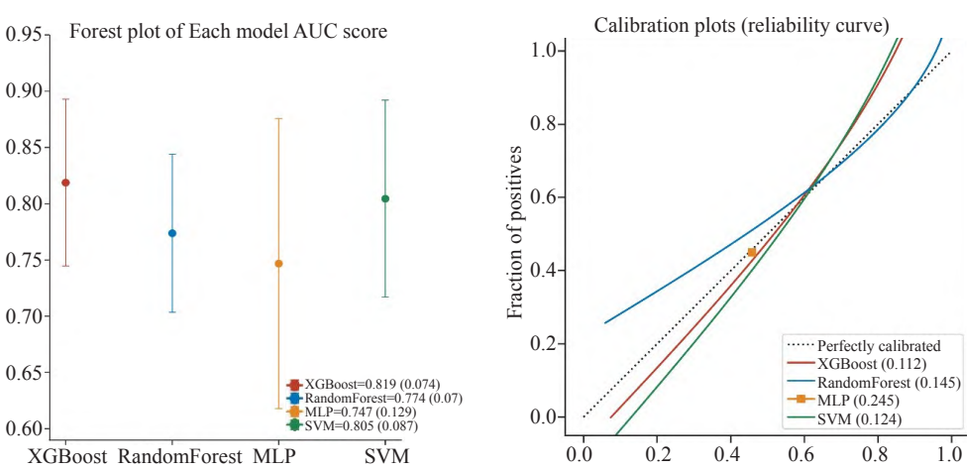

图6 验证集中4种ML模型生存预测的AUC得分森林图和校准曲线

Fig.6 AUC score forest plots and calibration curves for survival prediction using the 4 machine learning models in the validation set.

精准预测手术预后对AEG患者的后续治疗决策具有重要意义。目前,对于肿瘤术后的预后预测主要是基于Logistic回归模型和Cox回归建模[21,22],Logistic回归模型缺少生存时间,对生存预测不及Cox回归模型,本次研究使用了Cox-PH对AEG生存率进行建模预测,获得了较好的预测结果。Cox-PH的构建需要满足等比例假设,因此在构建该模型时,需要舍弃部分重要预后影响因素。Cox-PH属于线性回归,其预测需满足线性回归方程,无法捕捉特征之间的交互关系,相比之下,机器学习技术可以更好地捕捉到特征之间的复杂关联[23],从而提高模型的准确性。

本研究利用了性别、年龄、Borrmann分型、分化程度、浸润深度(T分期)、淋巴结转移数量(N分期)、病理TNM分期、肿瘤最大直径、术后化疗、Fibr、D-dimer、手术方式、术后住院时间、PNI、NLR、WBR、CEA、AFP、CA199共19项临床工作中易于获取的临床观察指标,构建了预测AEG 3年和5年生存状态的Cox-PH和4种机器学习模型。在最终建模的19项临床观察指标中,Fibr和D-dimer术前血液指标与胃癌生存率的相关性研究较少,故本次研究纳入以上两种临床观察指标,通过5种生存预测模型予以验证其与AEG预后具有相关性。Borrmann分型、分化程度、浸润深度(T分期)、淋巴结转

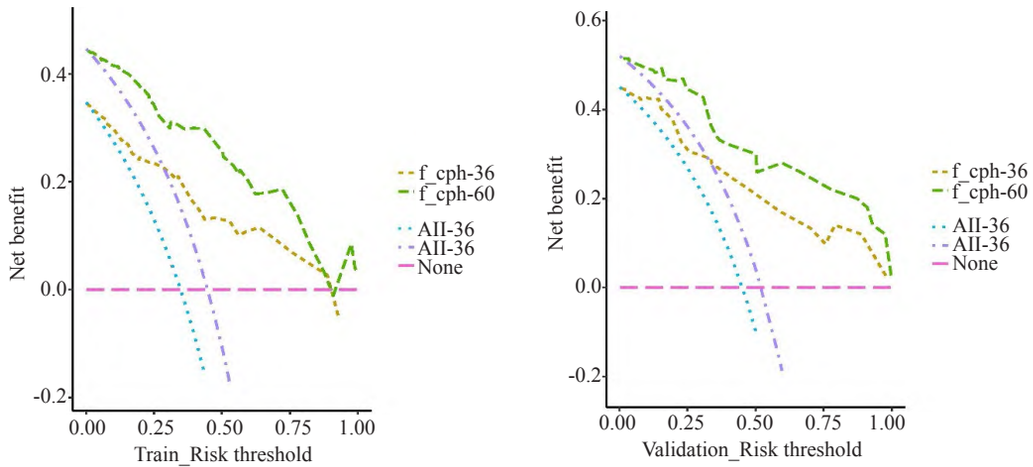

图7 Cox回归模型预测训练集和验证集的临床决策曲线(DCA)

Fig.7 Clinical decision curves (DCA) of the training and validation sets as predicted by the Cox regression model. With the threshold of the model set in the range of 10%-90%, the decision curve lies above the None line and All line, and thus the model is clinically useful.

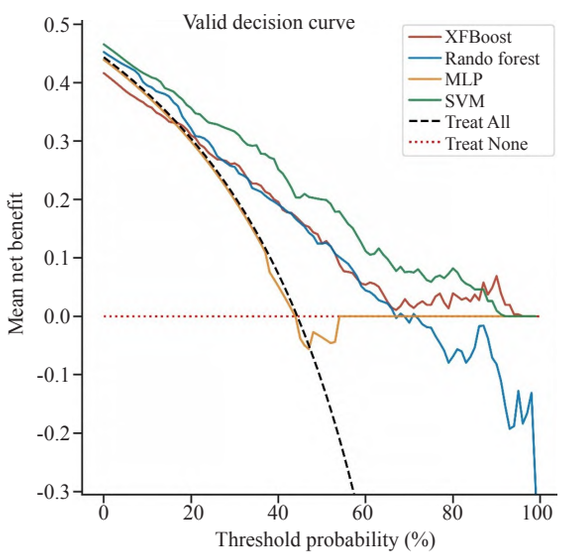

图8 4种ML模型预测的临床决策曲线(DCA)

Fig.8 Clinical decision curves (DCA) for predictions by the 4 machine learning models. The optimal cutoff values of XGBoost, RF, MLP, and SVM in ROC analysis of the validation set were 38.5% , 50.7% , 46.1% , and 46.0% , respectively. At these thresholds, the decision curves of XGBoost and SVM models lie above the None and All lines, while the decision curves of the MLP and RF models did not (or not completely) lie above the None and All lines.

移数量(N分期)、病理TNM分期、肿瘤最大直径是公认的影响AEG患者预后的因素[24-27]。浸润深度(T分期)、淋巴结转移数量(N分期)、病理TNM分期3种临床观察指标具有高度重叠,为防止机器学习模型的过拟合,本研究在构建5种模型时排除了浸润深度和淋巴结转移数量两种指标。在机器学习模型的构建中,对训练集应用交叉验证进行超参数调优,各模型均充分表达出其预测效能。综合各个模型在训练集和测试集ROC曲线的AUC值、校准曲线及DCA曲线表现,XGBoost模型在机器模型中表现最佳,其在训练集和测试集内的AUC值均≥0.80,Cox回归模型也具有较高的预测效能,但其算法的局限性和丢失重要临床观察指标的特性,使其不能与XGBoost相比。因此,本研究构建的XGBoost模型具有较高的实用性和可靠性。

本研究的局限性:本研究为单中心研究,纳入患者数量有限,机器学习模型在大数据集上应用可获得更加稳定的结果[28]。本次研究中MLP为深度学习模型,是ML模型的子集,为人工神经网络,具有较高的非线性特征数据学习模拟的能力,但在本研究中未表现出其应有的预测效能,可能原因是研究中,未有效提取变量特征,数据量较少。因此在后续的研究中,可加入多中心的数据进行训练和外部验证,从而得到更加可靠的预测模型。其次,家族史、吸烟、饮酒等可能影响AEG患者远期预后的因素未纳入本研究收集的19项临床观察指标,在后续的模型优化中,可加入更多可能影响AEG远期预后的因素,以不断完善预测模型。最后,本研究是利用回顾性数据开发和验证的,在正式临床实践前,还应进行前瞻性验证研究以确认模型的可靠性。

综上所述,本研究基于临床工作常见的19个临床病理特征构建了预测AEG术后生存风险的Cox-PH和机器学习模型,其中XGBoost模型效能最佳。可为AEG的预后评估、术后治疗决策提供重要参考,进而推动AEG的个体化诊治。

参考文献:

[1] Liu K, Yang K, Zhang W, et al. Changes of esophagogastric junctional adenocarcinoma and gastroesophageal reflux disease among surgical patients during 1988-2012: a single-institution, high-

volume experience in China[J]. Ann Surg, 2016, 263(1): 88-95.
[2] Imamura Y, Watanabe M, Toihata T, et al. Recent incidence trend of surgically resected esophagogastric junction adenocarcinoma and microsatellite instability status in Japanese patients[J]. Digestion, 2019, 99(1): 6-13.
[3] Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends[J]. Ann Oncol, 2012, 23(12): 3155-62.
[4] Kusano C, Gotoda T, Khor CJ, et al. Changing trends in the proportion of adenocarcinoma of the esophagogastric junction in a large tertiary referral center in Japan[J]. J Gastroenterol Hepatol, 2008, 23(11): 1662-5.
[5] Prentice RL, Zhao SS. Regression models and multivariate life tables[J]. J Am Stat Assoc, 2021, 116(535): 1330-45.
[6] Li LX. Dimension reduction for high-dimensional data[J]. Methods Mol Biol, 2010, 620: 417-34.
[7] Murphy K. Machine learning-a probabilistic perspective[M]. MIT Press, 2012.
[8] Verma AA, Murray J, Greiner R, et al. Implementing machine learning in medicine[J]. J De L'association Med Can, 2021, 193(34): E1351-7.
[9] Chip M, Lynch. Prediction of lung cancer patient survival *via* supervised machine learning classification techniques[J]. Int J Med Inform, 2017, 108: 1-8.
[10] Zhou CM, Xue Q, Wang Y, et al. Machine learning to predict the cancer-specific mortality of patients with primary non-metastatic invasive breast cancer[J]. Surg Today, 2021, 51(5): 756-63.
[11] Ji GW, Fan Y, Sun DW, et al. Machine learning to improve prognosis prediction of early hepatocellular carcinoma after surgical resection[J]. J Hepatocell Carcinoma, 2021, 8: 913-23.
[12] Kaitlin M, Christopherson, MD, et al. A machine learning model approach to risk-stratify patients with gastrointestinal cancer for hospitalization and mortality outcomes[J]. Int J Radiat Oncol, 2021, 111(1): 135-42.
[13] Nagtegaal Iris D, Odze Robert D, David K, et al. The 2019 WHO classification of tumours of the digestive system[J]. Histopathology, 2020, 76(2): 182-8.
[14] Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization[J]. Clin Cancer Res, 2004, 10(21): 7252-9.
[15] 赵 刚, 吴志勇, 卞正乾, 等. 贲门癌选择胸腹联合切口的标准[J]. 上海交通大学学报: 医学版, 2006, 26(8): 845-8.
[16] de Manzoni G, Pedrazzani C, Verlato G, et al. Comparison of old and new TNM systems for nodal staging in adenocarcinoma of the gastro-oesophageal junction[J]. Br J Surg, 2004, 91(3): 296-303.
[17] Tytgat GNJ, Bartelink H, Bernards R, et al. Cancer of the esophagus and gastric cardia: recent advances[J]. Dis Esophagus, 2004, 17(1): 10-26.
[18] Fein M, Fuchs KH, Ritter MP, et al. Application of the new classification for cancer of the cardia[J]. Surgery, 1998, 124(4): 707-13; discussion713-4.
[19] Kim S, Kim KO. Treatment of adenocarcinoma of the esophago-gastric junction[J]. Korean J Helicobact Upper Gastrointestinal Res, 2012,12(3): 151-9.
[20] 徐 宇, 朱蕙燕, 龙子雯, 等. 食管胃结合部腺癌的预后分析[J]. 中国癌症杂志, 2010, 20(6): 446-51.
[21] Liu XX, Guo W, Shi XB, et al. Construction and verification of prognostic nomogram for early-onset esophageal cancer[J]. Bosn J Basic Med Sci, 2021,21(6): 760-72.
[22] Li W, Xiao Y, Xu XW, et al. A novel nomogram and risk classification system predicting the cancer-specific mortality of patients with initially diagnosed metastatic cutaneous melanoma[J]. Ann Surg Oncol, 2021, 28(7): 3490-500.
[23] Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities[J]. Br J Gen Pract, 2018, 68(668): 143-4.
[24] Gao AQ, Wang LL, Li J, et al. Prognostic value of perineural invasion in esophageal and esophagogastric junction carcinoma: a meta-analysis[J]. Dis Markers, 2016, 2016: 7340180-91.
[25] Shahbaz Sarwar CM, Luketich JD, Landreneau RJ, et al. Esophageal cancer: an update[J]. Int J Surg, 2010, 8(6): 417-22.
[26] Yang J, Lu Z, Li L, et al. Relationship of lymphovascular invasion with lymph node metastasis and prognosis in superficial esophageal carcinoma: systematic review and meta-analysis[J]. BMC Cancer, 2020, 20(1): 176-89.
[27] Vaibhav, Gupta, MD, et al. Survival prediction tools for esophageal and gastroesophageal junction cancer: a systematic review[J]. J Thorac Cardiovasc Surg, 2018, 156(2): 847-56.
[28] van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints[J]. BMC Med Res Methodol, 2014, 14: 137-51.

（编辑：林 萍）