BMC Cancer

RESEARCH Open Access

Comparison of machine learning methods for Predicting 3-Year survival in elderly esophageal squamous cancer patients based on oxidative stress

Jin-Biao Xie[1*†], Shi-Jie Huang[1†], Tian-Bao Yang[1], Wu Wang[1], Bo-Yang Chen[1] and Lianyi Guo[2]

Summary

Background Oxidative stress process plays a key role in aging and cancer; however, currently, there is paucity of machine-learning model studies investigating the relationship between oxidative stress and prognosis of elderly patients with esophageal squamous cancer (ESCC).

Methods This study included elderly patients with ESCC who underwent curative ESCC resection surgery continuously from January 2013 to December 2020 and were stratified into the training and external validation cohorts. Using Cox stepwise regression analysis based on Akaike information criterion, the relationship between oxidative stress biomarkers and prognosis was explored, and a geriatric ESCC-related oxidative stress score (OSS) was constructed. To construct a predictive model for 3-year overall survival (OS), machine-learning strategies including decision tree (DT), random forest (RF), and support vector machine (SVM) were employed. These machine-learning strategies play a key role in data mining and pattern recognition tasks. Each model was tested in the external validation cohort through 1000 resampling iterations. Validation was conducted using receiver operating characteristic area under the curve (AUC) and calibration plots.

Results The training cohort and validation cohort consisted of 340 and 145 patients, respectively. In the training cohort, the 3-year OS rate for patients was 59.2%. We constructed the OSS based on systemic oxidative stress biomarkers using the training cohort. The study found that pathological N stage, pathological T stage, tumor histological type, lymphovascular invasion, CEA, OSS, CA 19 − 9, and the amount of bleeding were the most important factors influencing the 3-year OS. These eight important features were included in training the RF, DT, and SVM and trained on the training cohort and validated cohort, respectively. In the training cohort, the RF model demonstrated the highest predictive performance with an AUC of 0.975 (0.962–0.987), while the DT model is 0.784 (0.739–0.830) and the SVM is 0.879 (0.843–0.916). In the external validation cohort, the RF model again exhibited the highest

†Jin-Biao Xie and Shi-Jie Huang contributed equally to this work and should be considered co-first authors

*Correspondence:
Jin-Biao Xie
jinbiaoxie123@163.com

Full list of author information is available at the end of the article

performance with an AUC of 0.791 (0.717–0.864), compared to the DT model with an AUC of 0.717 (0.640–0.794) and 0.779 (0.702–0.856) in SVM.

Conclusions The random forest clinical prediction model constructed based on OSS can effectively predict the prognosis of elderly patients with ESCC after curative surgery.

Keywords Machine learning, Esophageal squamous cancer, Elderly patients, Oxidative stress, Prognosis

Introduction

Esophageal squamous cancer (ESSC) is one of the most common malignant tumors in the world with a high mortality rate [1]. Elderly patients with cancer demonstrate more variability in their physical, functional, psychological, and social strengths or vulnerabilities compared to younger patients [2]. Consequently, the current TNM staging system may not fully account for the specific prognostic features of elderly patients with esophageal squamous cell carcinoma. In recent years, studies have found that biological markers such as albumin (ALB), total bilirubin (TBIL), direct bilirubin (DBIL), blood urea nitrogen (BUN), creatinine (Crs), which reflect systemic oxidative stress, play important roles in the occurrence, development, and prognosis of elderly malignant tumor patients [3–6]. Antioxidant enzyme activity is believed to decline in elderly individuals, increasing their susceptibility to oxygen free radical damage. The increased susceptibility of elderly cancer patients' cells to oxidative stress can lead to damage of DNA, proteins, and lipids, speeding up cellular aging and disease progression [7–9]. Understanding the oxidative stress characteristics in elderly cancer patients can help doctors predict the complex prognosis of esophageal squamous cancer (ESCC), refine treatment plans, and improve the quality of life. Currently, the predictive models for survival in elderly patients with ESCC based on oxidative stress indicators remain lacking.

As part of artificial intelligence, supervised machine-learning techniques are widely employed to predict biological outcomes due to their ability to capture complex patterns, particularly in large and sparse datasets [10–12]. However, prior models depended on established variables such as TNM staging, histopathological characteristics, surgery, and chemotherapy, rendering them inadequate for the complex profiles of elderly individuals. Considering the significant impact of oxidative stress on elderly ESSC patients, this study aims to investigate the relationship between oxidative stress and the prognosis of elderly ESSC patients. Additionally, it seeks to develop a machine learning model to predict 3-year survival post-surgery, thereby aiding clinical decision-making.

Methods

Patient selection

This study included elderly ESCC patients who underwent curative ESCC resection surgery from January 2013 to December 2020 and were registered in the Thoracic Surgery Database of Putian University Affiliated Hospital (AHPTU) and Fujian Medical University Union Hospital (FMUUH). Inclusion criteria were as follows: (1) postoperative pathological diagnosis of Esophageal squamous cell carcinoma; (2) age $\geq$65 years at diagnosis; (3) underwent curative surgery with no evidence of distant metastasis; and (4) complete clinical and pathological data were available. Exclusion criteria were: (1) postoperative pathology confirmed a non-primary tumor originating from the esophagus; (2) presence of distant metastases; (3) incomplete clinical data. Ultimately, after applying exclusion criteria, 340 patients from AHPTU were included in the study group as the training cohort, and 145 patients from FMUUH were included in the external validation cohort. This study was a retrospective analysis of anonymized data from the database, and the Institutional Review Board waived the requirement for informed consent.

Validation selection

Other clinically relevant features used in training the machine learning predictive model were established by researchers based on clinical reasoning, literature review, and consensus on routine availability to ensure wide applicability in various clinical settings. Specifically, the predictive model included preoperative hematological tests (white blood cell count (WBC), hemoglobin (HB), neutrophils (NE), lymphocytes (LYM), monocytes (NOM), prothrombin time (PT)), biochemical tests((ALB), BUN, TBIL, DBIL, UA, Crs, LDH), tumor markers (alpha-fetoprotein (AFP), carcinoembryonic antigen (CEA), carbohydrate antigen 19−9 (CA199)), clinical variables (gender, age, OSS, body mass index, history of major abdominal surgery, history of previous malignancy, Charlson Comorbidity Index, American Society of Anesthesiologists (ASA) score, neoadjuvant therapy), intraoperative variables (extent of resection, intraoperative blood loss), postoperative variables (Clavien-Dindo complication grading, adjuvant chemoradiation), and pathological variables (differentiation grade, pathological T(pT) stage, pathological N(pN) stage, pathological TNM(pTNM) stage, lymphovascular invasion, perineural invasion). Due to potential collinearity among variables, other potential predictive factors were excluded from the candidate predictive variables (using pT stage, pN stage without using pTNM

stage). Variables were standardized to ensure comparability of scales.

Candidate predictive variables

Complete data on preoperative tests, intraoperative conditions, postoperative recovery, and pathological results are essential. Routine blood and biochemical tests were conducted for each patient from the first day of admission. The TNM staging was reclassified according to the American Joint Committee on Cancer/Union for International Cancer Control 8th edition AJCC Cancer Staging Manual. Based on the optimal cutoff values identified by R 4.3.3 software, each biochemical parameter was converted into a categorical variable. The oxidative stress indicators in our study included ALB, TBIL, DBIL, BUN, Crs, lactate dehydrogenase (LDH), and uric acid (UA). Based on the optimal cutoff values determined by the surv-cutpoint method, the biochemical parameters were classified as low (values below the cutoff) or high (values above the cutoff). The training cohort was used to develop a new Oxidative Stress Score (OSS) using beta coefficients from a multivariate stepwise Cox regression analysis. Patients were then stratified into risk groups based on the calculated optimal cutoff value of OSS, which was subsequently validated in the.

Establishment of machine learning model

To predict the survival status at 3 years post-surgery, we analyzed the discriminative capabilities of three classification machine learning algorithms: random forest (RF), decision tree (DT), and support vector machine (SVM). These methods were chosen due to their widespread application and superior performance in cohort studies. All statistical analyses were conducted using previously developed R packages: "randomForest," "MASS," "PRROC," "rpart," "caret," and "e1071." In order to select hyperparameters and the best probability, we trained the models using a cross validation scheme. DT is a supervised machine learning technique utilized for both regression and classification tasks. DT predicts the target variable's value by learning simple rules that are represented by a decision tree, comprising nodes, branches, and leaves. The algorithm classifies each sample by traversing the tree from the root to a leaf node. RF is an ensemble learning algorithm applicable to classification, regression, and unsupervised learning. It comprises multiple unpruned trees, each created using the DT algorithm through a recursive partitioning process. SVM is another widely used supervised learning algorithm for classification and regression. SVM constructs a hyperplane or multiple hyperplanes in a high-dimensional space, optimally separating data into different classes. For nonlinear classification, the Radial Basis Function kernel is used to estimate and maximize the hyperplane's margin.The training cohort was utilized to develop a novel oxidative stress score (OSS) and to construct a machine learning predictive model (Fig. 1).

Neoadjuvant/adjuvant chemoradiation

In accordance with the National Comprehensive Cancer Network guidelines [13], we recommend preoperative neoadjuvant radiochemotherapy for esophageal squamous cancer at stages cTis–2N1–3M0 or cT3–4aNany M0, and adjuvant chemotherapy was recommended for patients with suspicious cT4b or high-risk factors (T4a and N1–3 stages).The majority regimens comprised weekly carboplatin (area under curve 2) and paclitaxel (50 mg/m^2) for 5 weeks combined with daily radiotherapy consisting of 23 fractions of 1.8 Gy (total 41.4 Gy) [14]. After neoadjuvant therapy, surgery was conducted on patients who do not have distant metastasis and can generally tolerate surgery.

Follow-up

Follow-up visits were scheduled every three months during the first two years after surgery, and every six months from two to five years postoperatively. The final assessment took place in December 2023. Most routine

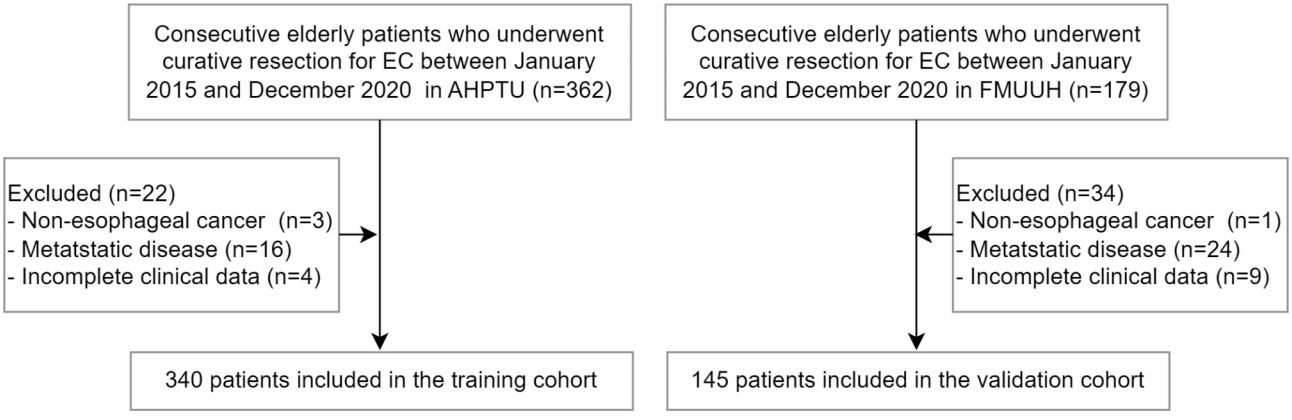

Fig. 1 Flow diagram of the study population

follow-up visits included physical examinations, laboratory tests, chest X-rays, abdominal ultrasounds or computed tomography scans, and annual endoscopic examinations. The primary outcome was defined as overall survival (OS) after discharge, which was measured from the date of surgery to the date of death from any cause or to the last follow-up date for censored observations. The 3-year follow-up rate is over 90%.

Statistical analysis

Data analysis was performed using R version 4.3.3. For continuous data not following a normal distribution, we applied the Mann-Whitney test, while the independent t-test was used for normally distributed continuous data. Differences in the distribution of categorical variables between groups were analyzed using Pearson's chi-squared test and Fisher's exact test. Overall survival (OS) curves were generated using the Kaplan-Meier method, with differences between survival curves assessed using the log-rank test. Validation was conducted via bootstrap resampling. Model parameters were trained using training cohorts, and the performance of the trained model was evaluated with independent validation datasets. The performance metrics for the trained classifier included sensitivity, specificity, accuracy, AUC values, and Brier scores. All statistical analyses were conducted using R software version 4.3.3 (https://www.r-project.org/), with a two-sided p-value of <0.05 considered statistically significant.

Results

Study cohort

A total of 485 elderly patients with ESCC were included in this study. In the training cohort, there were 340 patients, of whom 261 (76.8%) were males and 79 (23.2%) were females, with a median age of 69 (67–73) years. In the validation cohort, there were 145 patients, with 119 (82.1%) males and 26 (17.9%) females, with a median age of 69 (67–73) years old. There were statistical differences in the clinical and pathological data of patients in the training and validation cohorts in terms of pathological N staging, lympholymphovascular invasion, and complication grading ($p=0.007$, 0.047, 0.002, respectively), while other variables demonstrated no statistical differences (Table 1, $p>0.05$).

In terms of survival rates, the 3-year (OS) overall survival rate in the training cohort was 37.22% (29.78%, 46.52%), while in the validation cohort, it was 49.87% (39.50%, 62.97%) (eFigure 1). The 3-year recurrence rate in whole cohort was 36.4%.

Developing a novel oxidative stress score

In the training cohort, the optimal cutoff values for oxidative stress indicators were determined using the surv_cutpoint method as follows: ALB 39.4 g/dL, BUN 4.65 mg/dL, TBIL 5.7μmol/L, DBIL 1.5μmol/L, UA 215μmol/L, LDH 162 U/L, and Crs 49.3 μmol/L. These indicators were included in a Cox proportional hazards model to perform bidirectional stepwise regression method based on the Akaike information criterion. ALB, BUN, UA, LDH, and Crs were the core factors affecting the OS of elderly patients with ESCC in the final model (eTable 1). Proportional hazards assumptions were satisfied for all variables, as verified by the Schoenfeld residual plots (eFigure 2) and tests (eTable 2). Based on the regression coefficients of these variables, a prognostic model called the Esophageal Squamous Cancer Oxidative Stress Score (OSS) was further constructed as follows: OSS = ALB * −0.3197 + BUN * 0.2397 + UA * −0.4927 + LDH * 0.3392 + Crs * 0.6625. Patients were stratified into low and high-OSS groups using the optimal cutoff value of OSS (eFigure 3). Kaplan–Meier survival curve analysis indicated that the survival rate of patients in the low OSS group was significantly lower than that of the high OSS group ($p<0.05$). In the validation cohort, similar results were obtained (Fig. 2).

Variable selection

In our study, we used the Boruta algorithm to select important variables related to the disease status. Boruta is an important assessment method based on random forest (RF), which determines which variables are truly important by comparing the importance of the original variables with randomly generated "shadow" variables. After running 50 iterations, a set of variables such as pN, pT, tumor histology, lympholymphovascular invasion, CEA, OSS, CA 19−9, and bleeding were identified as important variables. The importance of these variables was visualized in graphical charts, where the importance of each variable is demonstrated by comparing its importance relative to the maximum importance of the shadow variables (eFigure 4). To comprehensively evaluate potential prognostic factors, we conducted univariate and multivariate Cox regression analyses on a wide range of clinical variables (eTable 3).

Model performance: validation

We included the selected variables in machine learning to construct three models (RF, DT, and SVM), and Table 2 reveal the performance metrics of these models in predicting 3-year OS in the training and validation cohorts. The area under the curve (AUC) for RF was 0.975 (0.962–0.987) and 0.791 (0.717–0.864); DT had AUC values of 0.784 (0.739–0.830) and 0.717 (0.640–0.794) in the training and validation cohorts, respectively; and SVM had AUC values of 0.879 (0.843–0.916) and 0.779 (0.702–0.856). Besides, the AUC of the Cox model is 0.625 (0.581–0.667) and 0.532 (0.510–0.553), respectively. Compared

Table 1 Comparison of clinicopathologic characteristics between patients included in the training and validation set

	level	Overall	Training	Validation	*p*
n		485	340	145	
Recurrence within 36 months		207	145	62	0.982
Dead within 36 months		235	166	69	0.803
Sex (%)	Female	105 (21.6)	79 (23.2)	26 (17.9)	0.239
	Male	380 (78.4)	261 (76.8)	119 (82.1)	
Age[median(QR)]		69(67–73)	69(67–73)	69(67–73)	0.904
Smoking history (%)	No	477 (98.4)	333 (97.9)	144 (99.3)	0.487
	Yes	8 (1.6)	7 (2.1)	1 (0.7)	
Alcohol consumption (%)	No	484 (99.8)	339 (99.7)	145 (100.0)	1.000
	Yes	1 (0.2)	1 (0.3)	0 (0.0)	
Charlson Comorbidity Index (%)	0	405 (83.5)	285 (83.8)	120 (82.8)	0.728
	1	62 (12.8)	43 (12.6)	19 (13.1)	
	2	2 (0.4)	2 (0.6)	0 (0.0)	
	3	15 (3.1)	9 (2.6)	6 (4.1)	
	5	1 (0.2)	1 (0.3)	0 (0.0)	
History of malignancy (%)	No	472 (97.3)	332 (97.6)	140 (96.6)	0.706
	Yes	13 (2.7)	8 (2.4)	5 (3.4)	
Surgical history (%)	No	431 (88.9)	302 (88.8)	129 (89.0)	1.000
	Yes	54 (11.1)	38 (11.2)	16 (11.0)	
BMI (mean (SD))		21.70 (3.01)	21.63 (3.03)	21.87 (2.97)	0.436
Differentation (%)	G1	47 (9.7)	35 (10.3)	12 (8.3)	0.895
	G2	183 (37.7)	129 (37.9)	54 (37.2)	
	G3	246 (50.7)	170 (50.0)	76 (52.4)	
	G4	9 (1.9)	6 (1.8)	3 (2.1)	
cT (%)	1	70 (14.4)	76 (22.4)	29 (20.0)	0.111
	2	60 (12.4)	44 (12.9)	18 (12.4)	
	3	48 (9.9)	30 (8.8)	10 (6.9)	
	4	307 (63.3)	190 (55.9)	88 (60.7)	
cN (%)	0	215 (44.3)	155 (45.6)	53 (36.6)	0.086
	1	85 (17.5)	47 (11.8)	32 (22.1)	
	2	97 (20.0)	58 (17.0)	25 (17.2)	
	3	88 (18.1)	80 (23.5)	35 (24.1)	
pT (%)	1	88 (18.1)	67 (19.7)	21 (14.5)	0.594
	2	50 (10.3)	34 (10.0)	16 (11.0)	
	3	20 (4.1)	14 (4.1)	6 (4.1)	
	4	327 (67.4)	225 (66.2)	102 (70.3)	
pN (%)	0	195 (40.2)	149 (43.8)	46 (31.7)	0.007
	1	73 (15.1)	40 (11.8)	33 (22.8)	
	2	87 (17.9)	62 (18.2)	25 (17.2)	
	3	130 (26.8)	89 (26.2)	41 (28.3)	
Lymphovascular invasion (%)	No	365 (75.3)	265 (77.9)	100 (69.0)	0.047
	Yes	120 (24.7)	75 (22.1)	45 (31.0)	
Perineural invasion (%)	No	350 (72.2)	253 (74.4)	97 (66.9)	0.114
	Yes	135 (27.8)	87 (25.6)	48 (33.1)	
Neoadjuvant chemoradiation (%)	No	417 (86.0)	289 (85.0)	128 (88.3)	0.419
	Yes	68 (14.0)	51 (15.0)	17 (11.7)	
Resection margin status (%)	R0	477 (98.4)	336 (98.8)	141 (97.2)	0.388
	R1-2	8 (1.6)	4 (1.2)	4 (2.8)	
Tumor histology (%)	Adenocarcinoma	197 (40.6)	144 (42.4)	53 (36.6)	0.276
	Squamous cell carcinoma	288 (59.4)	196 (57.6)	92 (63.4)	
Adjuvant chemoradiation(%)	No	208 (42.9)	153 (45.0)	55 (37.9)	0.18
	Yes	277 (57.1)	187 (55.0)	90 (62.1)	
Clavien Dindo Classification (%)	0	175 (36.1)	133 (39.1)	42 (29.0)	0.002

Table 1 (continued)

	level	Overall	Training	Validation	*p*
	<3	212 (43.7)	152 (44.7)	60 (41.4)	
	≥3	98 (20.2)	55 (16.2)	43 (29.7)	

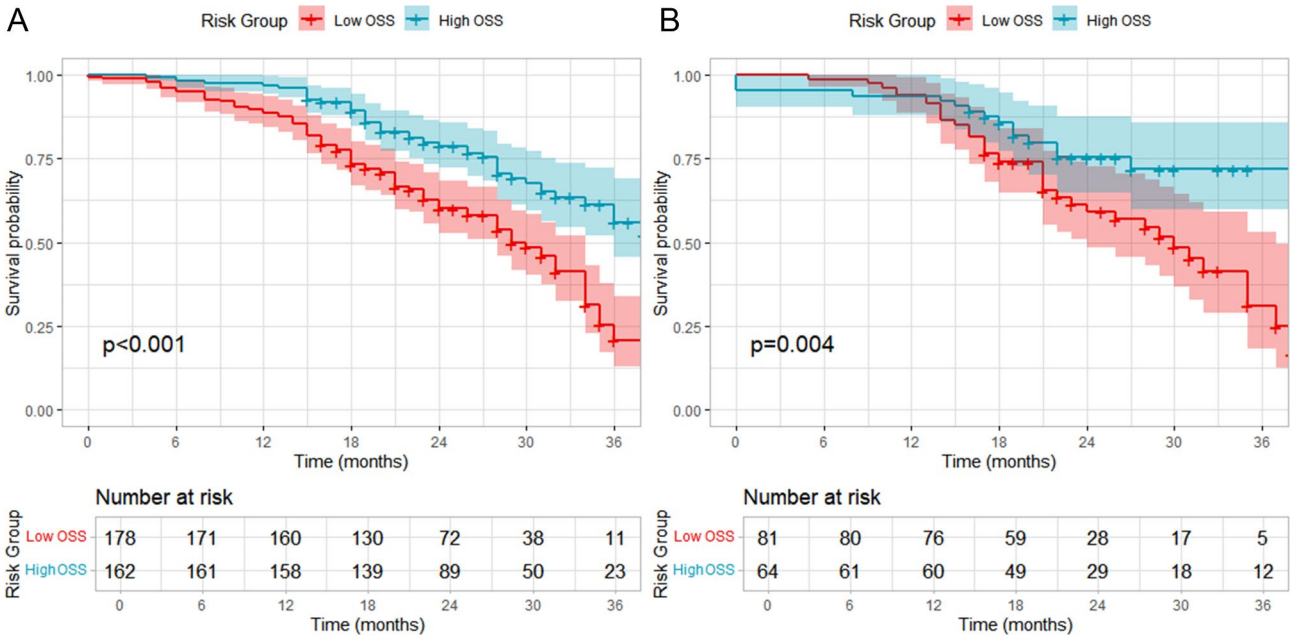

Fig. 2 Calibration curves of models
(**A**) Calibration curves of RSF model in training dataset; (**B**) Calibration curves of RSF model in validation dataset; (**C**) Calibration curves of DT model in training dataset; (**D**) Calibration curves of DT model in validation dataset; (**E**) Calibration curves of SVM model in training dataset; (**F**) Calibration curves of SVM model in validation dataset

Table 2 Classification performance of the individual model

		RF	DT	SVM
Training	AUC (95% CI)	0.975(0.962–0.987)	0.784(0.739–0.830)	0.879(0.843–0.916)
	Brier (95% CI)	0.075(0.065–0.086)	0.168(0.145–0.197)	0.143(0.124–0.161)
	Accuracy (95% CI)	0.903(0.872–0.934)	0.786(0.741–0.818)	0.803(0.756–0.843)
	Precision (95% CI)	0.885(0.828–0.926)	0.763(0.694–0.808)	0.816(0.74–0.87)
	Recall (95% CI)	0.943(0.904–0.982)	0.878(0.825–0.917)	0.822(0.772–0.869)
	F1 Score (95% CI)	0.913(0.888–0.943)	0.816(0.774–0.843)	0.819(0.77–0.855)
Validation	AUC (95% CI)	0.791(0.717–0.864)	0.717(0.640–0.794)	0.779(0.702–0.856)
	Brier (95% CI)	0.191(0.149–0.229)	0.218(0.185–0.266)	0.186(0.155–0.231)
	Accuracy (95% CI)	0.721(0.655–0.783)	0.684(0.614–0.766)	0.741(0.676-0.8)
	Precision (95% CI)	0.738(0.639–0.816)	0.675(0.587–0.765)	0.78(0.687–0.867)
	Recall (95% CI)	0.742(0.652–0.82)	0.79(0.712–0.868)	0.717(0.626–0.794)
	F1 Score (95% CI)	0.739(0.668–0.797)	0.727(0.66–0.801)	0.746(0.676–0.814)

to DT, SVM, and Cox model, RF had higher AUC values in the validation dataset, indicating that the RF model has excellent predictive performance and good generalization ability (see eFigure 5 A/B/C). We conducted an additional analysis comparing RF models with and without OSS. In the validation cohort, the AUC for the RF model without OSS was 0.7852. A DeLong test comparing these AUCs yielded a p-value of 0.7516.

In the training cohort, results from 1000 resamplings demonstrated that the Brier scores of RF, SVM, and DT were 0.075, 0.143, and 0.168, respectively. The Brier score is used to measure the accuracy of predicted probabilities, with lower scores indicating smaller deviations between predictions and actual outcomes, that is, higher prediction accuracy. These data indicate that in the training cohort, the RF model had the highest prediction accuracy, followed by SVM, while the DT model had a

relatively lower accuracy. In the resampling performance metrics on the test set, the Brier scores for RF, SVM, and DT were 0.191, 0.186, and 0.218, respectively. This highlights the advantage of RF and SVM over DT in terms of generalization ability (Table 2). Table 2 demonstrate the performance metrics of each model based on resampling.

Model performance: calibration and decision curve

The calibration curve plots demonstrate that the RF model performed well on all datasets, with predicted probabilities roughly matching the observed event frequencies. Particularly, in the validation cohort, the predicted probabilities were closer to the actual outcomes, indicating better generalization ability for the RF model (eFigure 6). Conversely, the DT model's predictions deviated significantly from the diagonal line on both datasets, indicating substantial differences between predicted and observed values in certain probability ranges, indicating poor calibration within these specific prediction probability intervals. In both the training and validation cohorts, the SVM model's calibration was close to the 45-degree line, with slight deviations in some probability intervals (Fig. 2), indicating good calibration of the SVM model in predicting probabilities and consistency across different datasets. Additionally, we used decision curve analysis to compare the clinical utility of these models. The results demonstrated that the RF model exhibited the highest net benefit in most threshold probability ranges in the training cohort, while the DT and SVM models had similar net benefits, but both were lower than the RF model (Fig. 3A). In the validation cohort, the RF model maintained high net benefits in most threshold probability ranges, while the DT and SVM demonstrate similar performance, both lower than the RF model (Fig. 3B).

Discussion

As the scholar believes that the impact of surgery on survival gradually diminishes after the third postoperative year, the 3-year survival rate following ESCC resection is a valuable audit indicator for evaluating the long-term quality of tumor surgical care [15–17]. Elderly individuals, as a special population, may face increased surgical risks due to their unique bio-psychosocial characteristics. These include a higher likelihood of complications, frailty, reduced stress tolerance, declining physical function, cognitive decline, and other factors that can complicate postoperative survival [18, 19]. In recent years, machine-learning methods have been widely employed to predict survival outcomes for patients with liver cancer, gastric cancer, colorectal cancer, breast cancer, and prostate cancer [20–24], demonstrating a strong predictive performance. In this study, we successfully predicted the 3-year survival rate after surgery for elderly patients with ESCC using machine-learning models (DT, RF, and SVM), with the RF model showing the best performance.

Previous clinical studies have preliminarily confirmed the predictive value of numerous clinicopathological biomarkers in forecasting recurrence, metastasis, and OS rates after ESCC surgery. These biomarkers include tumor size, venous invasion, differentiation status, and the TNM staging system [25–27]. However, these predictive biomarkers come with high detection costs and may not be

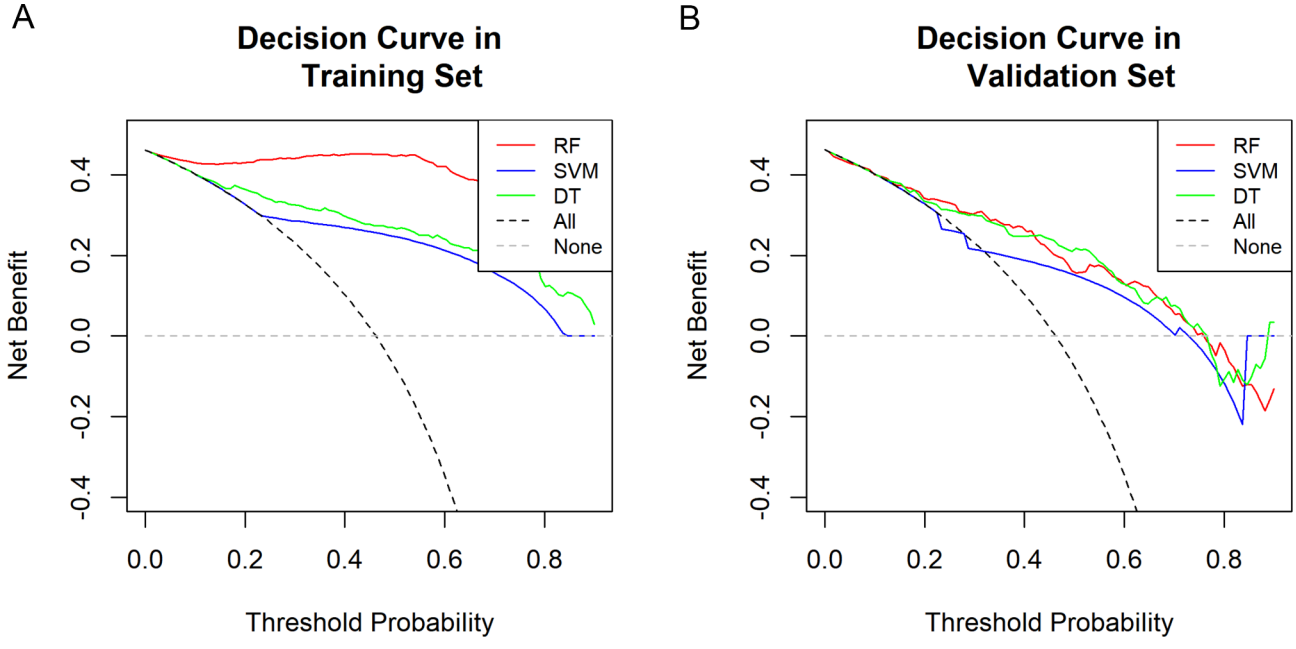

Fig. 3 Decision curve analysis curve of prediction models
Decision curve analysis curve in training dataset; (B) Decision curve analysis curve in validation dataset

suitable for a complex prognosis of elderly ESCC patients. Conversely, oxidative stress accelerates glycolysis, activates tumor cell migration, and promotes tumor proliferation. Furthermore, varying levels of oxidative stress can alter phosphorylation levels, influencing the malignancy and prognosis of tumors [28]. Some researchers suggest that oxidative stress might be linked to the expression of ferritin metabolism genes, thereby affecting prognosis [29]. Experiments using animal models have shown that, in response to external stimuli, mice exhibit increased oxidative stress factors, leading to significant elevations in biochemical markers such as TBIL, LDH, CRE, and BUN, thereby promoting tumor initiation and progression [30, 31]. Prospective studies have indicated that oxidative stress leads to alterations in patients' ALB, BUN, UA, LDH, and Cr levels. Following antioxidant therapy, patients demonstrated better scores, lower mortality rates, and decreased sepsis rates compared to the control group [32]. Although oxidative stress is linked to cancer, its use in predicting the prognosis of ESCC has not been extensively studied. To explore this, we proposed an oxidative stress index for ESCC, termed OSS, which includes ALB, DBIL, and BUN, based on preoperative hematological indicators associated with oxidative stress. Our study found that patients with lower OSS had a poorer prognosis compared to those with higher OSS. It is noteworthy that OSS was developed by training on a patient cohort from our institution, leveraging detailed clinical data and long-term follow-up. Thus, we hypothesize that a predictive model incorporating OSS and other pathological index may more accurately forecast the prognosis of elderly patients with ESCC.

Several models have been reported for predicting postoperative survival in ESCC. Li [33] built nomograms for predicting progression-free survival (PFS) and OS based on the Cox model to determine independent prognostic factors for PFS and OS. Following internal cross-validation, the corrected concordance indices were 0.739 and 0.696. However, inherent selection bias in retrospective studies and the inclusion of a limited number of cases further magnified this limitation. On the other hand, Xie [34] used a prospective study to construct an OS prediction model for patients with ESCC using LASSO regression. In the training and validation cohorts, the AUC was 0.811 (95% CI: 0.67–0.952) and 0.805 (95% CI: 0.638–0.973), respectively, indicating a strong predictive performance. However, this model did not specifically differentiate elderly ESCC patients. To determine whether the established model is applicable to elderly patients, further research is warranted. Contrarily, Liu [35] conducted a stratified analysis of survival characteristics in elderly patients with ESCC and developed a nomogram prognostic prediction model with a C-index of 0.706. This study utilized the SEER database as the validation cohort, resulting in the inclusion of fewer modeling indicators that may not comprehensively capture patient information. To address this limitation, Xie [36] explored the predictive value of basic indicators such as age, sex, and education level, including pathological indicators such as tumor staging, histology, and margin status, and surgical indicators such as neoadjuvant therapy, reoperation, and the Charlson comorbidity index for ESCC. Although this comprehensive model includes a wide range of information, it was established solely through multivariable regression analysis, lacking the generalizability and automation offered by machine learning, which might lead to the omission of key prognostic indicators. In response, we developed and validated various machine-learning methods (RF, DT, and SVM) to enhance the accuracy of predicting 3-year OS in elderly patients with ESCC. Compared to other models, the RF model showed excellent performance and good calibration in predicting the 3-year survival status. Our model utilized routine and easily obtainable perioperative clinical data, with key variables including pN, pT, tumor histological type, lymphovascular invasion, CEA, OSS, CA 19−9, and the amount of bleeding. Indeed, we observed the relationship between UA levels and prognosis, which may seem counterintuitive. However, studies have shown that UA may have antioxidant effects in certain circumstances, especially at lower concentrations, thereby exerting a protective effect on tumor cells [37]. Consequently, this approach provides a new opportunity to understand the significance of preoperative oxidative stress, patient status, surgical performance, postoperative recovery, and tumor staging in predicting the 3-year survival rate for elderly ESCC patients. This can aid clinicians improve the accuracy and effectiveness of patient management. Decision curve analysis allows us to evaluate and compare the performance of different models at various thresholds, aiding in model selection and application. Besides these, we noted that there were differences in the pathological and survival data between patients in the training and validation groups, yet our model still managed to show good predictive performance. This suggests that the RF model is applicable across different populations, exhibiting a strong generalization capability.

Our novel RF model demonstrated higher AUC values compared to previous models, likely attributed to the inclusion of more comprehensive clinical assessment indicators specific to elderly patients, such as oxidative stress markers, comorbidity indices, and complication status. While the model demonstrated excellent performance in the training cohort, its advantage over other models was less pronounced in the validation cohort. However, considering overall metrics including AUC, Brier score, and calibration curves, it still exhibited better performance and generalizability. Although pN and pT were

identified as the most decisive variables for model prediction, their importance remarkably surpassed that of other variables. Furthermore, tumor histological type, lymphovascular invasion, CEA, OSS, CA 19−9, and the amount of bleeding also revealed relatively high importance. These findings highlight the crucial variables in the model, aiding in its further optimization and interpretation. Thus, when developing a predictive model for long-term survival after ESCC surgery, these prognostic factors should be considered comprehensively.

Some limitations of this study should be acknowledged. As it was retrospective, selection bias cannot be completely avoided. The OSS was constructed by combining serum indicators such as ALB, BUN, UA, LDH, and Cr and may not fully capture the oxidative stress status of patients. A more accurate assessment of oxidative stress typically requires the detection of reactive oxygen species and their associated markers, such as superoxide dismutase and malondialdehyde. Furthermore, while the predictive model's performance was evaluated in terms of discriminative ability and risk calibration, and the self-resampling bootstrap method helped mitigate overfitting, the limited sample size still poses a risk of insufficient generalization. Furthermore, well-known factors influencing ESCC, such as high-risk genetic mutations, immunotherapy drug use, and socioeconomic status, were not available in our database and could potentially enhance model performance. To further validate these findings, future prospective studies are recommended.

Conclusion

The clinical predictive model for OSS constructed using machine-learning methods can effectively predict the prognosis of elderly patients with ESCC after curative surgery.

Abbreviations

ESCC	Esophageal squamous cancer
DT	Decision tree
RF	Random forest
SVM	Support vector machine
OS	Overall survival
AUC	Area under the curve
ALB	Albumin
TBIL	Total bilirubin
DBIL	Direct bilirubin
BUN	Blood urea nitrogen
Crs	Creatinine
WBC	White blood cell count
HB	Hemoglobin
NE	Neutrophils
LYM	Lymphocytes
NOM	Monocytes
PT	Prothrombin time

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-13115-7.

Supplementary Material 1

Author contributions

JX and JH wrote the main manuscript text. TY, WW, YC, LG analyzed the data. All authors reviewed the manuscript.

Funding

Science and Technology Foundation of Putian, Grant/Award (Number: 2023S3F005),the Health Science and Technology Foundation of Fujian Province, Grant/Award(Number: 2022QNA100).

Data availability

All data generated or analyzed during this study are included in the supplementary information files.

Declarations

Competing interests

The authors declare no competing interests.

Informed consent

Informed consent was obtained from each participant.

Human rights statement and informed consent

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and later versions. Informed consent or substitute for it was obtained from all patients for being included in the study.

Author details

[1]Department of Cardiothoracic Surgery, The Affiliated Hospital of Putian University, No.999 Dongzhen Road, Fujian 351100, China
[2]Department of Gastroenterology, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou 121001, China

Received: 17 July 2024 / Accepted: 28 October 2024
Published online: 21 November 2024

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019;144(8):1941–53.
2. Bron D, Soubeyran P, Fulop T. Innovative approach to older patients with malignant hemopathies. Haematologica. 2016;101(8):893–5.
3. Hayes J, Dinkova-Kostova A, Tew K. Oxidative stress in Cancer. Cancer Cell. 2020;38(2):167–97.
4. Arfin S, Jha N, Jha S et al. Oxidative stress in Cancer Cell Metabolism. Antioxidants Basel, Switzerland.2021;10(5).
5. Abdelhamid R, Nagano S. Crosstalk between oxidative stress and aging in Neurodegeneration disorders. Cells. 2023;12(5).
6. Wang J, Sun Y, Zhang X, et al. Oxidative stress activates NORAD expression by H3K27ac and promotes oxaliplatin resistance in gastric cancer by enhancing autophagy flux via targeting the miR-433-3p. Cell Death Dis. 2021;12(1):90.
7. Toh D, Lee W, Zhou H et al. Lycium barbarumWolfberry () consumption with a healthy Dietary Pattern lowers oxidative stress in Middle-aged and older adults: a Randomized Controlled Trial. Antioxidants (Basel, Switzerland). 2021;10(4).
8. Zhang J, Yang L, Xiang X, Li Z, Qu K, Li K. A panel of three oxidative stress-related genes predicts overall survival in ovarian cancer patients received platinum-based chemotherapy. Aging. 2018;10(6):1366–79.
9. Hoskin TS, Crowther JM, Cheung J, et al. Oxidative cross-linking of calprotectin occurs in vivo, altering its structure and susceptibility to proteolysis. Redox Biol. 2019;24:101202.
10. Deo R. Machine learning in Medicine. Circulation. 2015;132(20):1920–30.
11. Van Calster B, Wynants L. Machine learning in Medicine. N Engl J Med. 2019;380(26):2588.

12. Jiang Y, Zhang Z, Yuan Q, et al. Predicting peritoneal recurrence and disease-free survival from CT images in gastric cancer with multitask deep learning: a retrospective study. Lancet Digit Health. 2022;4(5):e340–50.
13. Ajani JA, D'Amico TA, Bentrem DJ, et al. Esophageal and Esophagogastric Junction Cancers, Version 2.2023, NCCN Clinical Practice guidelines in Oncology. J Natl Compr Canc Netw. 2023;21(4):393–422.
14. Shapiro J, van Lanschot JJB, Hulshof MCCM, van Hagen P, van Berge Henegouwen MI, Wijnhoven BPL, et al. CROSS study group. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer(CROSS): long-term results of a randomised controlled trial. Lancet Oncol. 2015;16:1090–8.
15. Stahl M, Walz MK, Riera-Knorrenschild J, et al. Preoperative chemotherapy versus chemoradiotherapy in locally advanced adenocarcinomas of the oesophagogastric junction (POET): long-term results of a controlled randomised trial. Eur J Cancer. 2017;81:183–90.
16. Ely S, Alabaster A, Dominguez DA, et al. Effect of thoracic surgery regionalization on 1- and 3-Year survival after Cancer Esophagectomy. Ann Surg. 2023;277(2):e305–12.
17. Erratum. Nivolumab Plus Cabozantinib vs Sunitinib for First-Line treatment of Advanced Renal Cell Carcinoma (aRCC): 3-Year Follow-Up from the phase 3 CheckMate 9 ER trial. J Clin Oncol. 2023;41(21):3767.
18. Bollschweiler E, Plum P, Mönig SP, Hölscher AH. Current and future treatment options for ESCC in the elderly. Expert Opin Pharmacother. 2017;18(10):1001–10.
19. Lu HW, Chen CC, Chen HH, Yeh HL. The clinical outcomes of elderly ESCC patients who received definitive chemoradiotherapy. J Chin Med Assoc. 2020;83(10):906–10.
20. Famularo S, Donadon M, Cipriani F, et al. Machine learning predictive model to Guide Treatment Allocation for recurrent Hepatocellular Carcinoma after surgery. JAMA Surg. 2023;158(2):192–202.
21. Lee K, Jang J, Yu Y, et al. Usefulness of artificial intelligence for predicting recurrence following surgery for pancreatic cancer: retrospective cohort study. Int J Surg (London England). 2021;93:106050.
22. van den Bosch T, Warps A, de Nerée Tot Babberich M, et al. Predictors of 30-Day mortality among Dutch patients undergoing colorectal Cancer surgery, 2011–2016. JAMA Netw open. 2021;4(4):e217737.
23. Clift A, Dodwell D, Lord S, et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. BMJ (Clinical Res ed). 2023;381:e073800.
24. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam V. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the Surveillance, Epidemiology, and end results (SEER) database. Lancet Digit Health. 2021;3(3):e158–65.
25. Powell AGMT, Eley C, Chin C et al. Prognostic significance of serum inflammatory markers in ESCC [published correction appears in Esophagus. 2021;18(3):710.
26. Xi Y, Lin Y, Guo W, et al. Multi-omic characterization of genome-wide abnormal DNA methylation reveals diagnostic and prognostic markers for esophageal squamous-cell carcinoma. Signal Transduct Target Ther. 2022;7(1):53.
27. Hagens E, Tukanova K, Jamel S, et al. Prognostic relevance of lymph node regression on survival in ESCC: a systematic review and meta-analysis. Dis Esophagus. 2022;35(1):doab021.
28. Liang J, Cao R, Wang X, et al. Mitochondrial PKM2 regulates oxidative stress-induced apoptosis by stabilizing Bcl2. Cell Res. 2017;27(3):329–51.
29. Wang X, Xu Y, Dai L, et al. A novel oxidative stress- and ferroptosis-related gene prognostic signature for distinguishing cold and hot tumors in colorectal cancer. Front Immunol. 2022;13:1043738.
30. Liu M, Rao H, Liu J, et al. The histone methyltransferase SETD2 modulates oxidative stress to attenuate experimental colitis. Redox Biol. 2021;43:102004.
31. Sandesc M, Rogobete A, Bedreag O, et al. Analysis of oxidative stress-related markers in critically ill polytrauma patients: an observational prospective single-center study. Bosnian J Basic Med Sci. 2018;18(2):191–7.
32. Lima W, Martins-Santos M, Chaves V. Uric acid as a modulator of glucose and lipid metabolism. Biochimie. 2015;116:17–23.
33. Li B, Wang R, Zhang T, et al. Development and validation of a nomogram prognostic model for ESCC patients with oligometastases. Sci Rep. 2020;10(1):11259.
34. Xie C, Yang P, Zhang X, et al. Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. EBioMedicine. 2019;44:289–97.
35. Liu X, Guo W, Shi X, et al. Construction and verification of prognostic nomogram for early-onset ESCC. Bosn J Basic Med Sci. 2021;21(6):760–72.
36. Xie SH, Santoni G, Mälberg K, Lagergren P, Lagergren J. Prediction model of long-term survival after ESCC surgery. Ann Surg. 2021;273(5):933–9.
37. Lin Y, Xie Y, Hao Z, et al. Protective effect of uric acid on ox-LDL-Induced HUVECs Injury via Keap1-Nrf2-ARE Pathway. J Immunol Res. 2021;2021:5151168.

Publisher's note