

Using Pathomics-Based Model for Predicting Positive Surgical Margins in Patients with Esophageal Squamous Cell Carcinoma: A Comparative Study of Decision Tree and Nomogram

Ze Tang*, Shiyun Feng*, Qing Liu, Yunze Ban, Yan Zhang

Department of Thoracic Surgery, The First Hospital of Jilin University, Changchun, 130021, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yan Zhang, Email zhangyan2016@jlu.edu.cn

Objective: Esophageal squamous cell carcinoma (ESCC) has a high incidence and mortality rate. Postoperative positive surgical margins (PSM) often correlate with poor prognosis. This study aims to develop and validate a predictive model for PSM positivity in ESCC patients, with the potential to guide preoperative planning and improve patient outcomes.

Methods: We conducted a retrospective analysis of 1776 patients who underwent esophageal cancer surgery at the First Affiliated Hospital of Jilin University between January 2015 and December 2023. Patients with visible residual tumors (R2) or microscopic residual tumors (R1) at the surgical margins were classified as having PSM. High-dimensional pathological features were extracted from digital pathological sections using CellProfiler software. The selected features were used to develop a predictive model based on decision trees and generalized linear regression, and the model was validated in an independent cohort. Clinically significant pathological factors ($P < 0.05$) were included in multivariate logistic regression for further validation. The model's performance was assessed using calibration curves and receiver operating characteristic (ROC) curves, generated with the Bootstrap method. Decision curve analysis (DCA) was employed to evaluate the clinical utility of the predictive model.

Results: A total of 229 patients (12.89%) were diagnosed with PSM. Logistic regression analysis identified multifocal lesions, vascular invasion, and pathomics-based features as independent predictors of PSM. The predictive model, represented by a decision tree, demonstrated good discrimination with an area under the ROC curve of 0.899 (95% CI: 0.842–0.956, $P < 0.001$), and a strong calibration curve between the predicted probability and the actual probability. Additionally, the nomogram demonstrated slightly inferior discrimination with an area under the ROC curve of 0.803 (95% CI: 0.734–0.872, $P < 0.001$) in the training cohort.

Conclusion: Our study successfully established and validated a pathology-based predictive model for PSM risk, which could enhance preoperative evaluation and inform treatment strategies for ESCC.

Keywords: esophageal squamous cell carcinoma, surgical margin, machine learning, pathomics, prediction model

Introduction

Esophageal cancer is the eighth most common malignancy worldwide and the sixth leading cause of cancer-related deaths.^{1,2} More than half of global cases occur in China, where esophageal squamous cell carcinoma (ESCC) accounts for over 90% of diagnoses.³ The majority of ESCC patients present locally advanced disease at the time of diagnosis, and surgery is typically the preferred treatment method.⁴ It is worth mentioning that PSMs after esophagectomy are associated with a poor prognosis, including lower survival rates and higher recurrence risks due to residual tumors at the proximal, distal, or circumferential margins.^{4,5}

Nowadays, esophagectomy remains a cornerstone of comprehensive treatment.^{6,7} While advancements in surgical techniques have reduced the incidence of complications and perioperative mortality, the focus has shifted towards achieving high-quality oncological outcomes. Ensuring adequate tumor-free margins and effective lymph node dissection are critical for achieving radical resection.^{8,9} Meanwhile, avoiding PSM is also a key indicator of surgical quality. Previous studies have demonstrated that factors influencing the risk of PSM include tumor location, T3 or higher stage tumors, malnutrition, and preoperative radiotherapy and chemotherapy.^{8,10} Meanwhile, early prediction and identification models for esophageal cancer are gradually being developed and clinically tested.^{11–13} However, there is still a need for simple and effective tools to predict surgical margin status preoperatively, which could aid surgeons in planning and adjusting treatment strategies to reduce the incidence of residual tumors while ensuring effective resection.

Currently, visual observation remains the primary method for assessing pathological sections. Notably, the development of high-throughput processing technology for medical images has led to the emergence of “pathomics”, which involves the extraction of quantitative features from digital pathology images.¹⁴ Pathomic features can provide valuable insights into the tumor microenvironment, and recent research has shown their potential in cancer risk stratification, prognosis prediction, and adjuvant chemotherapy efficacy prediction.^{15,16} Encouraged by this, we developed a machine learning-based model using pathomic features to predict the risk of postoperative PSM in ESCC patients. This model could assist clinicians in making informed decisions and promoting individualized treatment plans. Given this situation, this study aims to retrospectively analyze the current status of PSM in esophagectomy and establish a clinically feasible prediction model to assess the likelihood of PSM in ESCC patients, followed by its validation.

Materials and Methods

Study Population

Retrospective analysis of 1776 patients who underwent surgical treatment for esophageal squamous cell carcinoma in the esophageal cancer comprehensive management database of the First Affiliated Hospital of Jilin University from January 2015 to December 2023. Inclusion criteria: (1) Postoperative pathological diagnosis of ESCC; (2) Age 18 years old and above; (3) The tumor is located in the thoracic esophagus. Exclusion criteria: (1) Postoperative pathology is non squamous cell carcinoma; (2) Cervical or esophagogastric junction cancer; (3) Patients undergoing salvage surgery; (4) Patients with incomplete clinical data. This study adopted the 8th edition of the esophageal cancer TNM staging system jointly released by the Joint Committee on Cancer of the United States and the International Union Against Cancer, which was officially implemented in 2018. This project has been approved for implementation by the Medical Ethics Committee of the First Affiliated Hospital of Jilin University (NO.20240812). All patients participating in this study have signed a letter of consent and strictly follow the Helsinki Declaration. We make sure that the patients were informed about the purpose of the study. The process of patient enrollment and the construction of prediction models are summarized in [Figure 1](#).

Definition of PSM

According to the College of American Pathologists criteria, R1 resection is defined as the presence of cancer cells at any point along the longitudinal (proximal or distal) or circumferential margins. Visible tumor residue detectable by the naked eye characterizes R2 resection.^{17,18} Patients with either R1 or R2 resection were classified as having PSM for this study.

Establishment of Pathological Omics Scoring System

We developed a case omics signature based on a survival analysis framework for whole section histopathological images using conventional HE staining. Specifically, this process includes four main stages: (1) generating candidate patches from pathological images of the entire sliced tissue; Set a fixed area sampling rate to sample candidate patches from WSI [patch size is 512×512 , 0.5 microns per pixel]; (2) Perform phenotype-based clustering on candidate patches; To distinguish patches from different parts (tumor/normal/both), cluster them based on their phenotypes. Due to the consideration of the high dimensionality of generated features, PCA was used for dimensionality reduction before

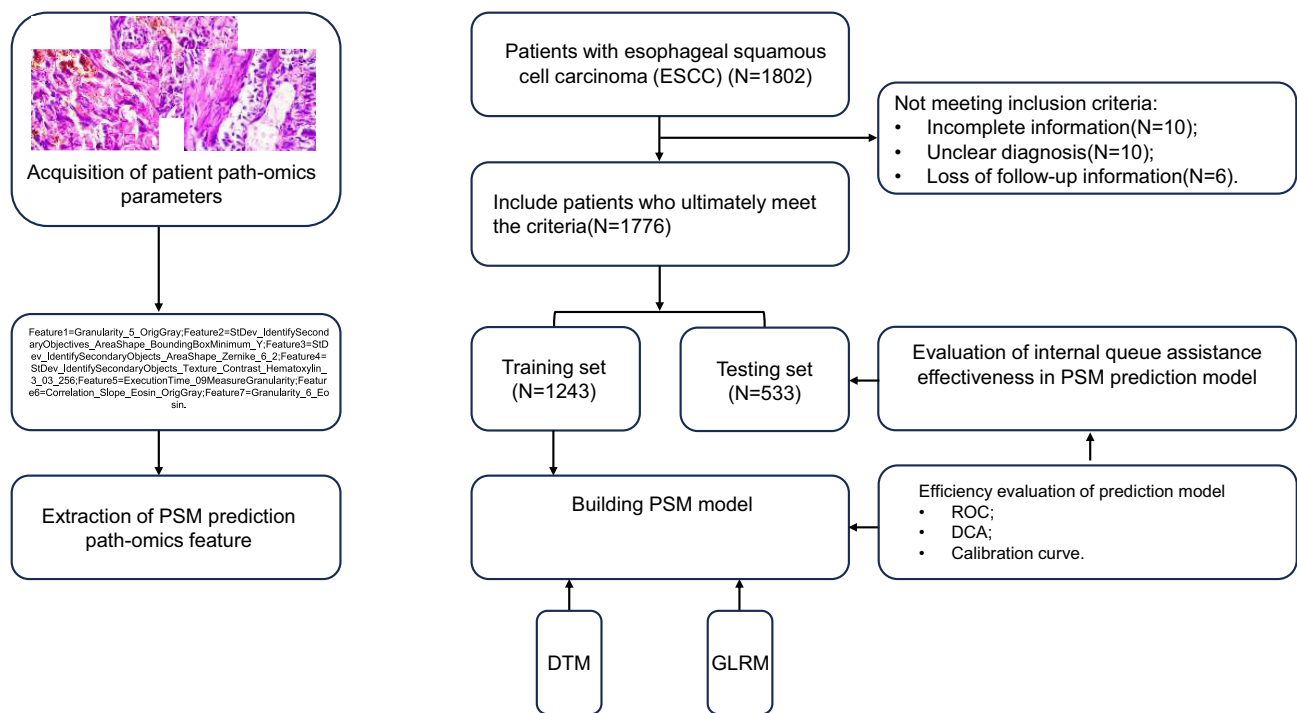


Figure 1 Process diagram for obtaining clinical pathological data of patients and constructing PSM prediction model. (A) Construction and extraction of digital pathological features; (B) Construction and efficiency evaluation of PSM prediction model.

implementing the K-means clustering process; (3) Identify target category clusters based on patch-wise survival prediction performance; Execute patch-wise training on a candidate subset and select a combination of candidate parameter categories with better prediction accuracy than random guessing.

Pathological Omics Feature Extraction

Quantitative features from the selected pathological image were extracted using the image analysis software CellProfiler (version 4.0.7).¹⁹ Initially, the “Unmix Colors” module separated the original stained images, converting them into grayscale images stained with hematoxylin and eosin. Additionally, the “ColorToGray” module was employed to convert the original stained images into grayscale.²⁰ The measurement process was divided into overall measurement and object measurement. In the overall measurement phase, 40–130 pathological features were extracted from each image. For object measurement, the hematoxylin-stained images were used to identify primary and secondary objects, followed by feature measurement. The average, median, and standard deviation of raw values from many objects within each image were calculated, resulting in the extraction of pathological features.

Construction of PSM Prediction Model

To remove redundant features, Mann–Whitney *U*-test was performed on each feature with a P-value of 0.05. Subsequently, considering the correlation between features, Spearman correlation analysis was performed on the features. If the absolute value of the correlation coefficient between two features is greater than 0.9, one of the features was excluded. Then, use the minimum absolute value convergence and selection operator algorithm to select the extracted features, and use 5-fold cross validation to select Lambda values to determine the optimal feature subset. Based on the selected optimal features, a pathological omics prediction model is constructed using decision tree and generalized linear regression algorithms. The optimal regularization parameters *C* and *Gamma* for the Gaussian radial basis function kernel are determined through 5-fold cross validation and grid search. Use the predicted output value of the pathological omics model as the risk coefficient for PSM occurrence.

Statistical Analysis

We used Jupyter notebook and R-studio for statistical analysis. Use chi-square test to compare whether there are significant differences in clinical indicators and pathological parameters between the training set and the test set; Kolmogorov Smirnov test was used to test the normality of the age of two groups of patients in the training and validation sets. Independent sample *t*-test was used for those who met the normal distribution, expressed in the form of mean \pm standard deviation; Otherwise, Mann–Whitney *U*-test will be used to compare whether there is a statistically significant difference between the two groups, represented by MD (P25, P75). The area under the curve (AUC) of the receiver operating characteristic curve (ROC) is used to evaluate the discriminative performance of the pathological omics model; Multi-factor logistic regression is suitable for further validation of independent predictive factors. Test level $\alpha = 0.05$.

Results

Baseline Characteristics and Pathological Parameters of Esophageal Cancer Patients

As shown in [Table 1](#) and [Supplementary Table 1](#), a total of 229 patients (12.89%) with PSM were included, comprising 149(65.07%) patients who underwent R1 resection and 80(34.93%) patients who underwent R2 resection. The lower segment of the esophagus was the most common site of onset. A subgroup analysis of the upper thoracic region revealed no statistically significant difference in the anastomotic site between neck and chest surgical margins. In addition, approximately 18.8% of R0 resection patients underwent laparoscopic surgery, while 81.2% of PSM patients underwent open surgery. Notably, 89.1% of patients with multifocal lesions were found to have PSM after surgery.

Selection of Candidate Parameters for the PSM Prediction Model

We first conducted a correlation analysis between the candidate predictor variables and the outcome variable (ie PSM), and the results showed that VI, MI, preoperative treatment (PT) were significantly correlated with PSM ([Figure 2](#)). So, we used LASSO analysis to obtain the optimal combination variables, that is, set the minimum penalty coefficient as the cutoff value, and included VI, MI, and some pathological parameters (feature 1, feature 3, and feature 6) for weight ranking and distribution. The results showed that VI, MI, and pathological parameters occupied the top weight values among the candidate predictive variables. Meanwhile, multivariate logistic regression analyses identified MI (OR=1.19,95% CI: 0.62–3.26,P < 0.05), VI(OR=1.61,95% CI: 1.04–2.99,P < 0.01) and pathomics-based features as statistically significant factors ([Table 2](#)).

Table 1 Clinical Baseline and Histopathological Data of Patients with Esophageal Squamous Cell Carcinoma

| Variables | Overall(N=1776) | Yes(N=229) | No(N=1547) | P-value |
|---------------------------------------|----------------------|----------------------|----------------------|---------|
| Sex (%) | | | | |
| Male | 875 (49.3) | 117 (51.1) | 758 (49.0) | 0.603 |
| Female | 901 (50.7) | 112 (48.9) | 789 (51.0) | |
| Age (median [IQR]), year | 48.00 [37.00, 60.00] | 47.00 [38.00, 59.00] | 48.00 [36.00, 60.00] | 0.915 |
| KPS (%), score | | | | |
| <90 | 922 (51.9) | 122 (53.3) | 800 (51.7) | 0.711 |
| ≥90 | 854 (48.1) | 107 (46.7) | 747 (48.3) | |
| BMI (median [IQR]), kg/m ² | 23.22 [17.72, 29.16] | 23.16 [17.48, 29.64] | 23.24 [17.87, 29.12] | 0.841 |
| Tumor location (%) | | | | |
| Upper | 578 (32.5) | 61 (26.6) | 517 (33.4) | 0.123 |
| Middle | 596 (33.6) | 83 (36.2) | 513 (33.2) | |
| Lower | 602 (33.9) | 85 (37.1) | 517 (33.4) | |
| T stage (%) | | | | |
| T1–T2 | 880 (49.5) | 118 (51.5) | 762 (49.3) | 0.568 |

(Continued)

Table I (Continued).

| Variables | Overall(N=1776) | Yes(N=229) | No(N=1547) | P-value |
|-------------------------|----------------------|----------------------|----------------------|---------|
| T3~T4 | 896 (50.5) | 111 (48.5) | 785 (50.7) | |
| N stage (%) | | | | 0.795 |
| N0~N1 | 866 (48.8) | 114 (49.8) | 752 (48.6) | |
| N2~N3 | 910 (51.2) | 115 (50.2) | 795 (51.4) | |
| TNM (%) | | | | 0.125 |
| I~II | 936 (52.7) | 132 (57.6) | 804 (52.0) | |
| III~IV | 840 (47.3) | 97 (42.4) | 743 (48.0) | |
| NLYM (median [IQR]) | 64.00 [40.00, 87.00] | 61.00 [41.00, 83.00] | 65.00 [40.00, 88.00] | 0.186 |
| VI (%) | | | | <0.001 |
| Yes | 552 (31.1) | 173 (75.5) | 379 (24.5) | |
| No | 1224 (68.9) | 56 (24.5) | 1168 (75.5) | |
| PI (%) | | | | 0.326 |
| Yes | 872 (49.1) | 105 (45.9) | 767 (49.6) | |
| No | 904 (50.9) | 124 (54.1) | 780 (50.4) | |
| MI (%) | | | | <0.001 |
| Yes | 818 (46.1) | 204 (89.1) | 614 (39.7) | |
| No | 958 (53.9) | 25 (10.9) | 933 (60.3) | |
| Surgery (%) | | | | <0.001 |
| Open | 301 (16.9) | 186 (81.2) | 115 (7.4) | |
| Laparoscope | 1475 (83.1) | 43 (18.8) | 1432 (92.6) | |
| AS (%) | | | | 0.293 |
| Neck | 899 (50.6) | 108 (47.2) | 791 (51.1) | |
| Chest or abdomen | 877 (49.4) | 121 (52.8) | 756 (48.9) | |
| PT (%) | | | | <0.001 |
| Yes | 1071 (60.3) | 39 (17.0) | 1032 (66.7) | |
| No | 705 (39.7) | 190 (83.0) | 515 (33.3) | |
| Feature1 (median [IQR]) | 3.82 [3.06, 4.64] | 8.54 [7.08, 10.04] | 3.64 [2.96, 4.31] | <0.001 |
| Feature2 (median [IQR]) | 2.95 [2.32, 3.62] | 1.42 [1.17, 1.64] | 3.13 [2.56, 3.72] | <0.001 |
| Feature3 (median [IQR]) | 5.83 [4.62, 7.07] | 2.38 [1.80, 3.03] | 6.15 [5.09, 7.29] | <0.001 |
| Feature4 (median [IQR]) | 3.03 [2.60, 3.52] | 3.06 [2.60, 3.50] | 3.03 [2.59, 3.52] | 0.828 |
| Feature5 (median [IQR]) | 8.14 [7.07, 9.23] | 30.64 [21.88, 39.55] | 7.84 [6.96, 8.79] | <0.001 |
| Feature6 (median [IQR]) | 3.14 [1.99, 4.16] | 0.73 [0.31, 1.13] | 3.42 [2.39, 4.30] | <0.001 |
| Feature7 (median [IQR]) | 43.42 [32.83, 55.25] | 43.65 [32.78, 54.97] | 43.42 [32.94, 55.31] | 0.995 |

Notes: Feature1, Granularity_5_OrigGray; Feature2, StDev_IdentifySecondaryObjectives_AreaShape_BoundingBoxMinimum_Y; Feature3, StDev_IdentifySecondaryObjectives_AreaShape_Zernike_6_2; Feature4, StDev_IdentifySecondaryObjectives_Texture_Contrast_Hematoxylin_3_03_256; Feature5, ExecutionTime_09MeasureGranularity; Feature6, Correlation_Slope_Eosin_OrigGray; Feature7, Granularity_6_Eosin.

Abbreviations: IQR, Interquartile range; KPS, Karnofsky Performance Status Score; BMI, Body Mass Index; TNM, Tumor node metastasis classification; NLYM, Number of lymph node dissections; VI, Vascular invasion; PI, Perineural invasion; MI, Multifocal lesions; AS, Anastomotic site; PT, Preoperative treatment.

Establishment and Validation of a Model for Predicting Positive Surgical Margins

A predictive model was constructed based on independent influencing factors for PSM and visualized in a Nomogram (Figure 3A). Scores corresponding to each predictive indicator were obtained from the chart, and their sum represented the total score, which predicted the risk of postoperative PSM occurrence. The decision tree model, as shown in Figure 3B, first divided cases based on whether the feature6 exceeded 1.5. In short, clinicians can determined the next risk value based on the characteristic parameter threshold of each node, until the final risk coefficient of the patient's PSM is obtained (Figure 4). DCA demonstrated that the decision tree model had the highest net benefit evaluation, consistent with its AUC value, whereas the generalized linear regression model showed the least benefit (Figure 5). As for the evaluation of predictive performance, the decision tree model showed an AUC value of 0.899 (95% CI: 0.0.842–0.956) in the training set and 0.852 (95% CI: 0.844–0.958) in the test set, while the generalized linear regression

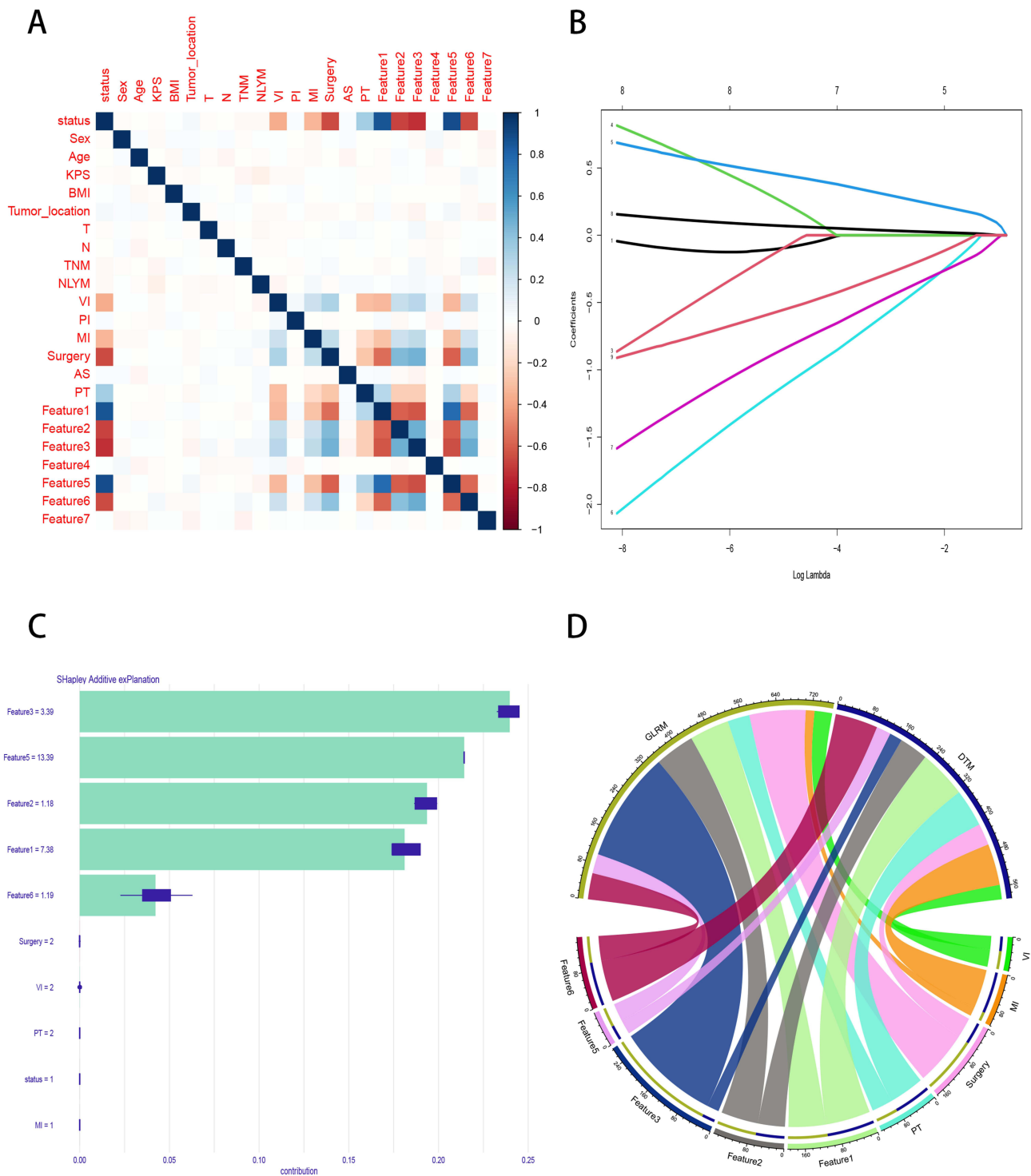


Figure 2 Selection of candidate variables for PSM prediction model. **(A)** The correlation analysis between PSM and candidate variables, where the darker the blue color intersecting the horizontal and vertical axes, the stronger the positive correlation, and vice versa; **(B)** Obtaining the best candidate combination variables based on the LASSO optimal penalty coefficient; **(C)** Weight ranking of optimal combination variables based on SHapley Additive exPlanation; **(D)** The allocation proportion of candidate variable weights in GLRM and DTM prediction models is represented by the width of the chordal graph dispersion, where the wider the width, the larger the weight value, and vice versa.

model had AUC values of 0.803 (95% CI: 0.734–0.872) and 0.793 (95% CI: 0.724–0.862) in the training and test sets, respectively (Table 3). Overall, the machine learning-based risk prediction model for PSM demonstrated satisfactory robustness, with the decision tree algorithm achieving optimal prediction performance.

Table 2 Logistic Regression Analysis of Pathology-Based Factors for PSM in Patients with ESCC

| Variables | Univariate analysis | | P-value | Multivariate analysis | | P-value |
|-----------|---------------------|-----------|---------|-----------------------|-----------|---------|
| | OR | 95% CI | | OR | 95% CI | |
| MI | 1.18 | 0.61~3.75 | <0.05 | 1.19 | 0.62~3.26 | <0.05 |
| VI | 1.56 | 0.99~3.13 | <0.05 | 1.61 | 1.04~2.99 | <0.01 |
| Feature1 | 4.26 | 2.69~4.83 | <0.01 | 3.99 | 1.42~4.56 | <0.01 |
| Feature3 | 1.95 | 0.38~2.52 | <0.01 | 1.87 | 0.30~2.44 | <0.01 |
| Feature6 | 2.26 | 0.69~2.83 | <0.01 | 2.38 | 0.71~2.85 | <0.01 |

Notes: Feature1, Granularity_5_OrigGray; Feature3, StDev_IdentifySecondaryObjects_AreaShape_Zernike_6_2; Feature6, Correlation_Slope_Eosin_OrigGray.

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval; MI, Multifocal lesions; VI, Vascular invasion.

Comparison Between Machine Learning Model for Predicting PSM and Manual Recognition

The decision tree model was compared with the evaluation level of senior pathology chief physicians. Encouragingly, [Figure 6](#) indicated that the decision tree model effectively identified high-risk PSM, matching the differential diagnostic ability of senior chief physicians. Overall, the machine learning model we developed for predicting PSM risk demonstrated high robustness and precision learning curves, achieving diagnostic capabilities comparable to those of experienced senior chief physicians.

Discussion

Positive margins after esophagectomy are linked to higher rates of recurrence and mortality.^{21,22} Achieving negative margins during esophagectomy for optimizing patient outcomes.^{23,24} Advances in imaging, preoperative neoadjuvant therapy, and intraoperative frozen section analysis have improved the ability to achieve negative margins.²⁵ This study analyzed clinical data from a large cohort of patients undergoing esophagectomy at a single center and developed a simple predictive model that can assist clinicians in preoperative evaluation and treatment planning. The study found that 12.89% of patients had non-R0 resections, a rate consistent with findings from the National Cancer Database (NCDB) in the United States, where positive margin after esophagectomy exceeds 9%.²⁶ This rate is higher compared to other cancer surgeries, such as colorectal and lung cancer, indicating the complexity of achieving negative margins in esophageal cancer. Additionally, we found that approximately 18.8% of R0 resection patients underwent laparoscopic surgery, while 81.2% of PSM patients underwent open surgery. We speculate that positive margins after esophagectomy often indicate incomplete tumor removal, necessitating subsequent adjuvant therapy. Herein, identifying factors that contribute to positive margins and establishing a predictive model for clinical use can significantly improve patient prognosis and reduce their economic and psychological burdens.

Our logistic regression analysis identified multifocal lesions and vascular invasion as risk factors for PSMs, while lower segment tumors, more extensive lymph node dissection, and laparoscopic surgery were also threatening factors in other literature reports.^{27,28} In this study, multifocal lesions were regarded as statistically independent risk factors for PSM, and most patients with multifocal tumors did have PSM post-surgery. This suggests that these patients may benefit from more extensive preoperative treatment or wider resection. Other studies have identified similar risk factors, including male gender, tumor length, T4 stage, and the Ivor-Lewis esophagectomy procedure.²⁹⁻³¹ While our study did not find preoperative treatment to be an independent factor for PSM, this may be due to biases in our data, particularly the low proportion of neoadjuvant therapy before 2018. Previous studies have shown that neoadjuvant therapy can significantly increase the likelihood of complete resection, contributing to improved long-term outcomes.^{32,33} Previous studies also found a correlation between the number of lymph nodes dissected and the likelihood of positive margins.^{34,35} Patients with more than 12 dissections were less likely to have tumor residues, suggesting that extensive

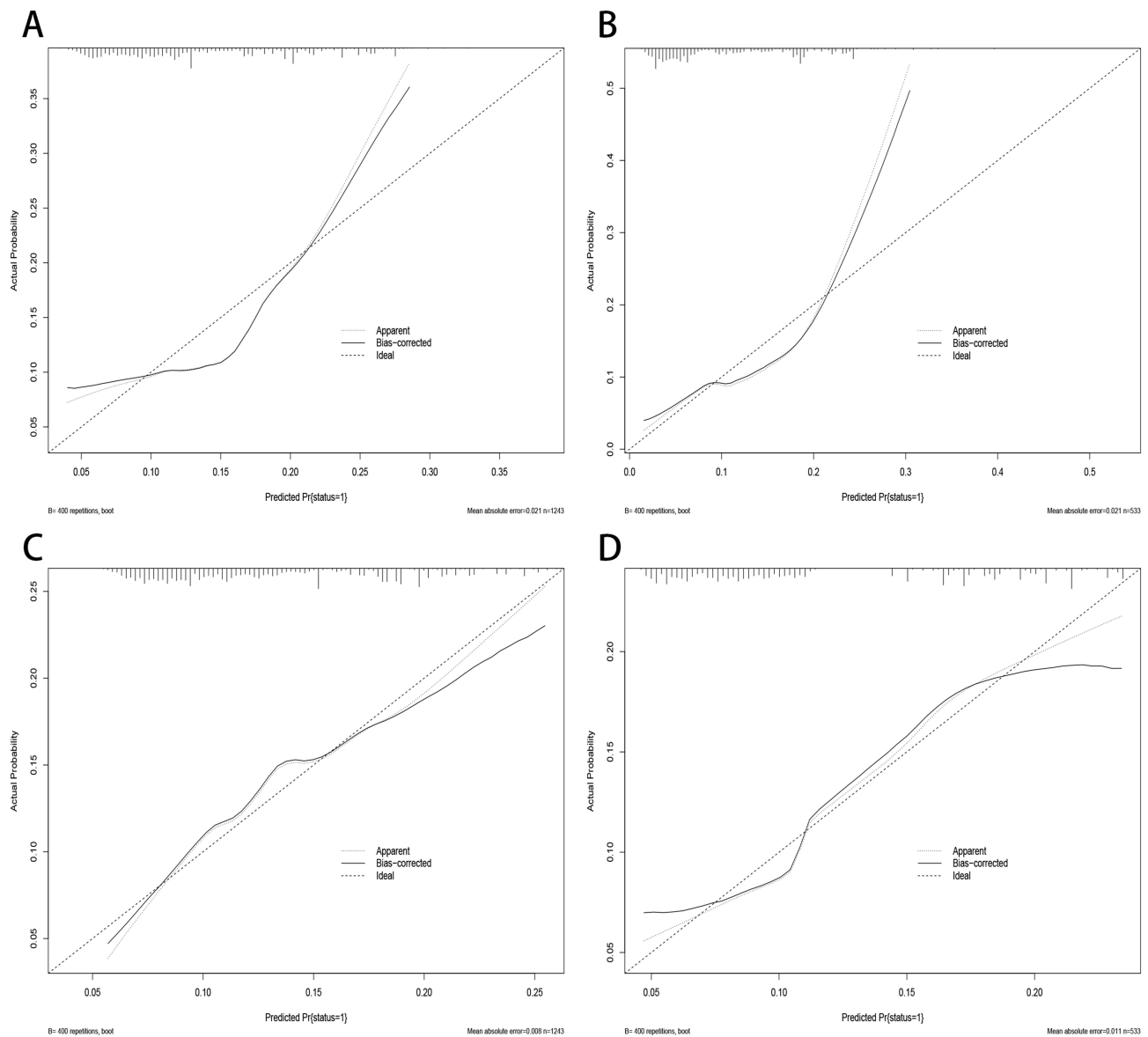


Figure 4 Stability evaluation of two prediction models. Correction curve evaluation of GLRM prediction model in (A) training set and (B) validation set; Evaluation of calibration curves for decision tree prediction models in (C) training set and (D) validation set.

lymph node dissection may reflect a more rigorous surgical approach, reducing the risk of PSMs. Currently, the 8th edition of the National Comprehensive Cancer Network Guidelines recommends clearing at least 15 lymph nodes. The thoroughness of lymph node dissection reflects the surgeon's commitment to achieving negative margins, as a less rigorous approach can increase the risk of tumor residue.

In clinical practice, the decision to perform rapid frozen section pathological analysis during surgery is often based on the surgeon's experience or specific intraoperative conditions. However, this empirical and situational decision-making can lack precision, potentially leading to PSM at the resection site. Fortunately, pathological genomics has emerged as a valuable tool for studying tumor cell heterogeneity and predicting tumor prognosis. By identifying relevant spatial relationships to classify cell interactions and signal transduction, as well as quantifying the intrinsic variability of different phenotypes and biological behaviors in tumor cells, this approach helps analyze and predict clinical outcomes and treatment responses following tumor surgery. In this study, we extracted a large number of pathological features from H&E-stained slides using CellProfiler image analysis software and applied the LASSO regression algorithm to propose specific pathological features. The optimal cutoff value was determined to be 1.12 through a maximum selection rank

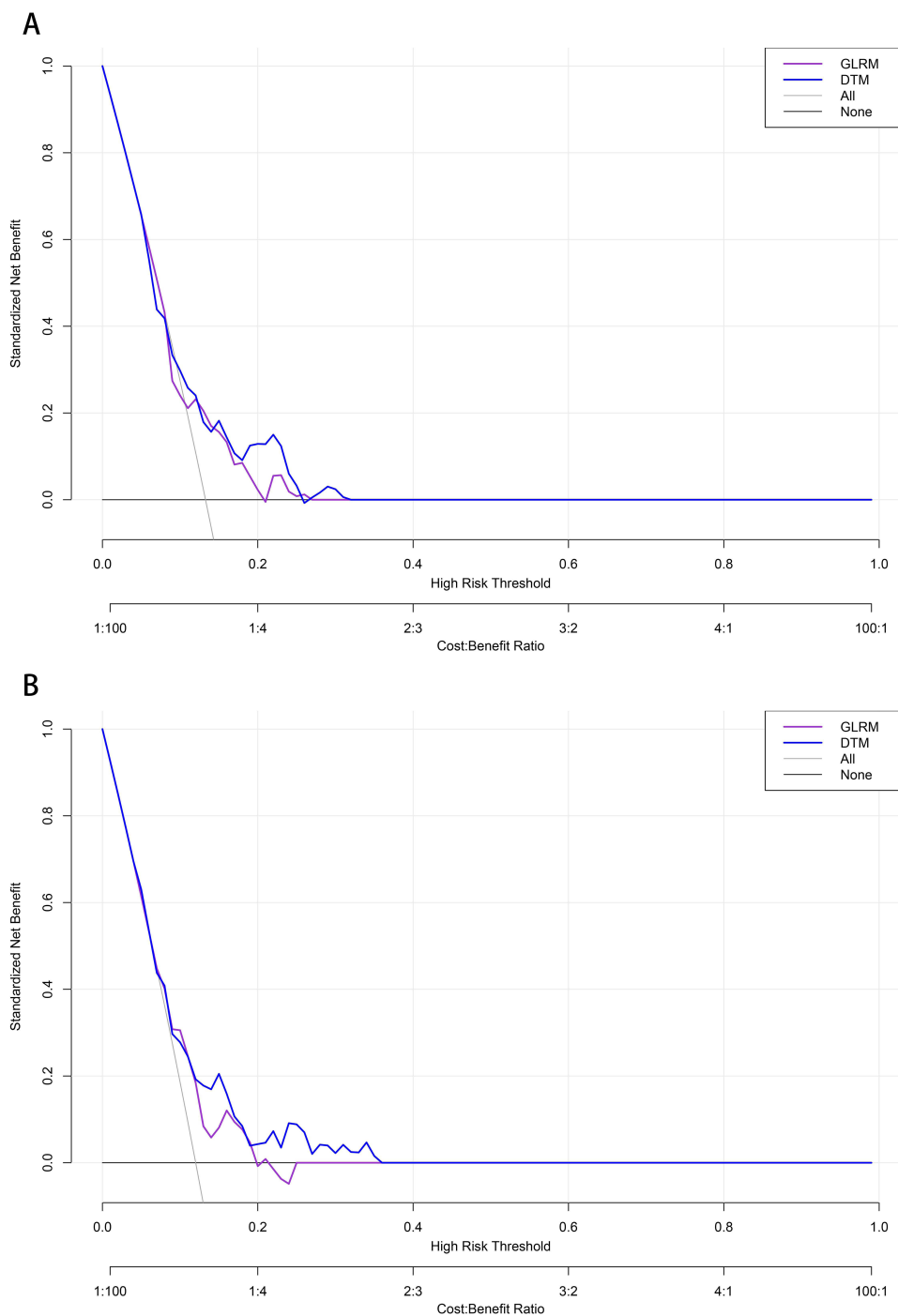


Figure 5 Net profit analysis of prediction model based on decision curve. **(A)** Training cohort; **(B)** Testing cohort.

Notes: The gray line represents the threshold, and the blue and purple lines represent DTM and GLRM, respectively. The farther the line is from the threshold, the greater the net benefit, indicating that the predictive model has better predictive performance.

test, and it was confirmed that this score is associated with high risk and prognosis of esophageal squamous cell carcinoma surgical margins. These findings suggest that pathological feature scores may serve as a potential biomarker for predicting surgical margins and prognosis in ESCC.

The predictive model developed in this study may assist clinicians in identifying patients at high risk for PSM, allowing for proactive adjustments in treatment plans and surgical strategies. For high-risk patients, more intensive preoperative treatment may be considered to reduce the risk of incomplete resection. Preoperative treatment can

Table 3 Evaluation of Predictive Performance of PSM Prediction Model Based on ROC

| Prediction model | Training set | | | | International set | | | |
|------------------|--------------|-------------|-------|-------|-------------------|-------------|-------|-------|
| | AUC | 95% CI | PPV | NPV | AUC | 95% CI | PPV | NPV |
| DTM | 0.899 | 0.842–0.956 | 0.991 | 0.852 | 0.901 | 0.844–0.958 | 0.992 | 0.895 |
| GLRM | 0.803 | 0.734–0.872 | 0.846 | 0.713 | 0.793 | 0.724–0.862 | 0.829 | 0.706 |

Abbreviations: AUC, Area under the curve; 95% CI, 95% confidence interval; PPV, Positive predictive value; NPV, negative predictive value; DTM, Decision tree model; GLRM, Generalized linear regression.

transform unresectable tumors into resectable ones, downstage locally advanced tumors, and maximize the likelihood of complete resection. This is particularly important for cases involving multifocal lesions, TNM staging, and individualized neoadjuvant therapy or conversion therapy should be strongly considered. If feasible, rapid frozen section pathological analysis during surgery to determine whether to re-perform resection or expanded organ resection, ensuring the thorough removal of the tumor. Surgeons must continually enhance their surgical expertise and accumulate experience to ensure accuracy and completeness in their procedures. The improvement of surgical techniques and surgical instruments and equipment can also enhance the accuracy and visibility of surgical resection, helping to reduce the occurrence of positive surgical margins after surgery. Advancements in surgical techniques, instruments, and equipment can also improve the precision of resections, helping to minimize the occurrence of positive surgical margins. In summary, the predictive tool from this study can aid clinicians in planning surgical strategies, such as expanding resection margins or incorporating intraoperative frozen section analysis, and in adjusting treatment decisions by adding or intensifying neoadjuvant therapy, thereby maximizing the chances of achieving complete resection.

This study also found that, compared to the R0 resection group, the PSM group exhibited a higher proportion of lymph node positivity but underwent fewer lymph node dissections. This disparity may be because of extensive lymph node metastasis, which often correlates with a broader tumor infiltration range, making surgery more challenging. To ensure patient safety, surgeons may limit the extent of resection, which underscores the importance of precise pre-operative evaluation and reducing tumor burden. This study attempted to predict the likelihood of a PSM before surgery to guide the use of intraoperative frozen sections. However, due to the limitations of retrospective data, some post-operative information was used to assist in decision-making.

There are several limitations to this study. Firstly, it is a single-center retrospective analysis conducted over an extended period, making it difficult to maintain consistency in surgical quality; Additionally, most patients with esophageal cancer are already locally advanced at the time of initial treatment, and neoadjuvant therapy has become the standard treatment recommended by various guidelines. Unfortunately, because the first version of the esophageal cancer diagnosis and treatment guidelines was only released by the Chinese Society of Clinical Oncology in 2019, the proportion of patients receiving neoadjuvant therapy before 2018 was extremely low, limiting further discussion on this topic. Moreover, this study only broadly explored factors influencing surgical margins without separately analyzing the proximal, distal, and circumferential margins after esophageal cancer surgery. This was due to the long duration of the study and the gradual improvement of pathological reporting during the construction of the single esophageal cancer full-process management database. More detailed analyses will be conducted in the future. Finally, as this study relied on data from retrospective research databases, it was not possible to accurately collect and record tumor recurrence for related analysis. Moving forward, we plan to expand the sample size of the population research cohort, develop multi-center prospective PSM prediction models based on multi-omics, and conduct external validation to improve the accuracy and the generalizability of these models.

Conclusion

PSM remains a significant factor influencing postoperative treatment decisions and patient prognosis. This study, utilizing a large retrospective dataset developed a predictive model to assess the risk of PSM in patients with ESCC. The model shows promise for application in preoperative evaluation and treatment planning for esophageal cancer patients. Despite advances in medical technology leading to a gradual decrease in the incidence of PSM, it remains a challenging clinical issue. Surgeons must leverage pathological omics information and predictive tools to further reduce

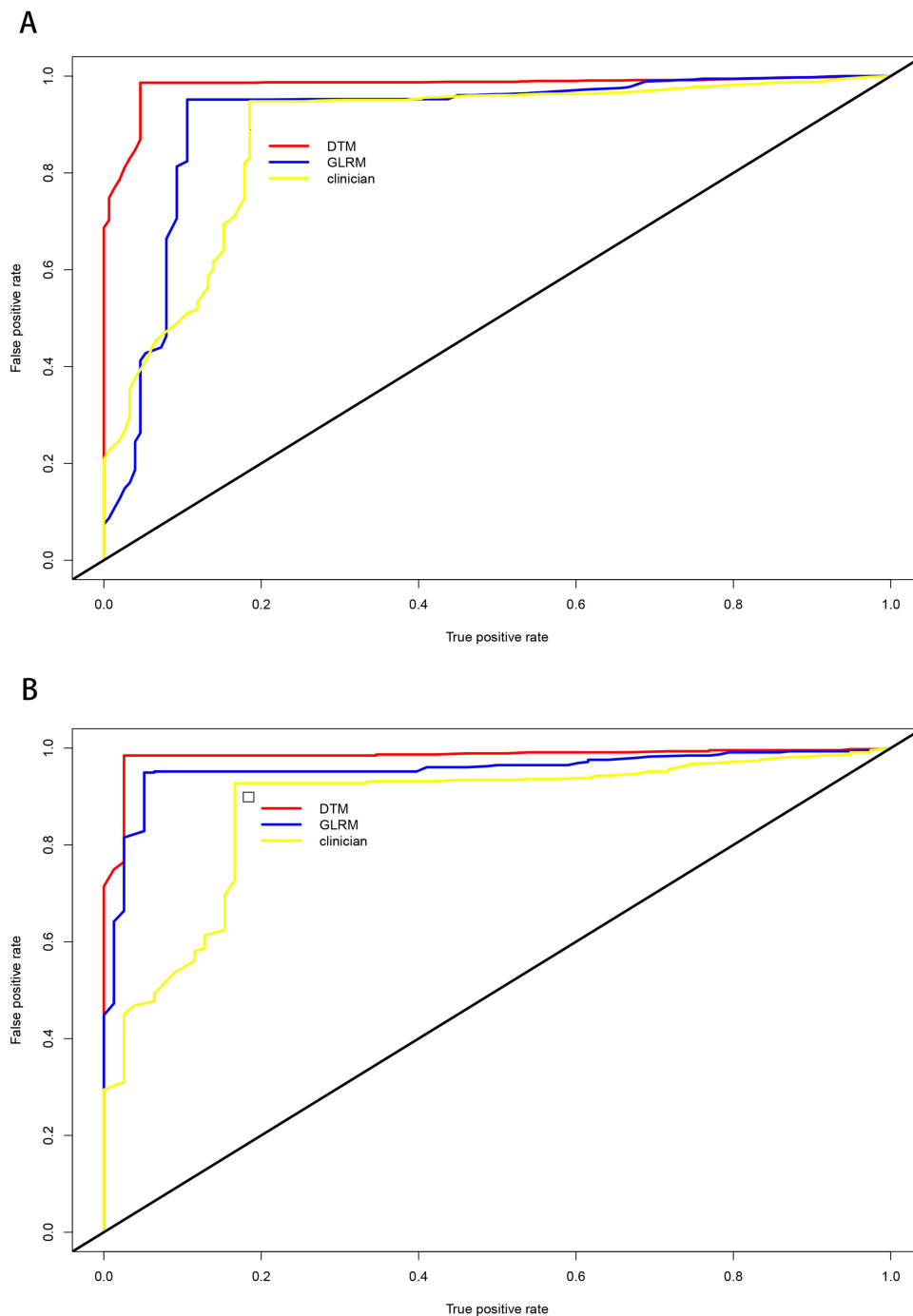


Figure 6 Evaluation of predictive model based on the area under the receiver working characteristic curve and the predictive performance of senior chief physicians. **(A)** Training cohort; **(B)** Testing cohort.

PSM occurrence. Building on the factors and models explored in this study, future prospective clinical trials are essential for both internal and external validation. Such studies will better equip clinicians to make informed preoperative preparations and surgical plans, ultimately reducing the risk of PSM and offering long-term benefits to a broader patient population.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Lam AK. Introduction: esophageal squamous cell carcinoma-current status and future advances. *Meth Mol Biol.* 2020;2129:1–6.
2. Abnet CC, Arnold M, Wei WQ. Epidemiology of esophageal squamous cell carcinoma. *Gastroenterology.* 2018;154(2):360–373. doi:10.1053/j.gastro.2017.08.023
3. Morgan E, Soerjomataram I, Rumgay H, et al. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: new estimates from GLOBOCAN 2020. *Gastroenterology.* 2022;163(3):649–58.e2. doi:10.1053/j.gastro.2022.05.054
4. Waters JK, Reznik SI. Update on management of squamous cell esophageal cancer. *Current Oncol Rep.* 2022;24(3):375–385. doi:10.1007/s11912-021-01153-4
5. Feng Y, Wei W, Guo S, Li BQ. Additional endoscopic treatments for patients with positive lateral margins after endoscopic resection of early esophageal squamous cell carcinoma. *Oncol Lett.* 2023;25(2):67. doi:10.3892/ol.2022.13653
6. Mayanagi S, Haneda R, Inoue M, Ishii K, Tsubosa Y. Selective lymphadenectomy for salvage esophagectomy in patients with esophageal squamous cell carcinoma. *Ann Surg Oncol.* 2022;29(8):4863–4870. doi:10.1245/s10434-022-11625-x
7. Di Pietro M, Canto MI, Fitzgerald RC. Endoscopic management of early adenocarcinoma and squamous cell carcinoma of the esophagus: screening, diagnosis, and therapy. *Gastroenterology.* 2018;154(2):421–436. doi:10.1053/j.gastro.2017.07.041
8. Feng Y, Wei W, Guo S, Li BQ. Associated risk factor analysis and the prognostic impact of positive resection margins after endoscopic resection in early esophageal squamous cell carcinoma. *Exp Ther Med.* 2022;24(1):457. doi:10.3892/etm.2022.11384
9. Hsu PK, Lee YY, Chuang LC, Wu YC. Lymph node dissection for esophageal squamous cell carcinoma. *Thoracic Surg Clin.* 2022;32(4):497–510. doi:10.1016/j.thorsurg.2022.07.001
10. Wen J, Linghu E, Yang Y, et al. Relevant risk factors and prognostic impact of positive resection margins after endoscopic submucosal dissection of superficial esophageal squamous cell neoplasia. *Surg Endosc.* 2014;28(5):1653–1659. doi:10.1007/s00464-013-3366-9
11. Chou CK, Karmakar R, Tsao YM, et al. Evaluation of spectrum-aided visual enhancer (SAVE) in esophageal cancer detection using YOLO frameworks. *Diagnostics.* 2024;14(11):1129. doi:10.3390/diagnostics14111129
12. Fang YJ, Huang CW, Karmakar R, et al. Assessment of narrow-band imaging algorithm for video capsule endoscopy based on decorrelated color space for esophageal cancer: part II, detection and classification of esophageal cancer. *Cancers.* 2024;16(3). doi:10.3390/cancers16030572
13. Yang KY, Mukundan A, Tsao YM, Shi XH, Huang CW, Wang HC. Assessment of hyperspectral imaging and CycleGAN-simulated narrowband techniques to detect early esophageal cancer. *Sci Rep.* 2023;13(1):20502. doi:10.1038/s41598-023-47833-y
14. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. *Med Image Anal.* 2016;33:170–175. doi:10.1016/j.media.2016.06.037
15. Chen D, Fu M, Chi L, et al. Prognostic and predictive value of a pathomics signature in gastric cancer. *Nat Commun.* 2022;13(1):6903. doi:10.1038/s41467-022-34703-w
16. Fogo AB. Learning from deep learning and pathomics. *Kidney Int.* 2023;104(6):1050–1053. doi:10.1016/j.kint.2023.06.006
17. Barbour AP, Rizk NP, Gonen M, et al. Adenocarcinoma of the gastroesophageal junction: influence of esophageal resection margin and operative approach on outcome. *Ann Surg.* 2007;246(1):1–8. doi:10.1097/01.sla.0000255563.65157.d2
18. Miner TJ, Jaques DP, Karpeh MS, Brennan MF. Defining palliative surgery in patients receiving noncurative resections for gastric cancer. *J Am Coll Surg.* 2004;198(6):1013–1021. doi:10.1016/j.jamcollsurg.2004.02.007
19. Stirling DR, Carpenter AE, Cimini BA. CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. *Bioinformatics.* 2021;37(21):3992–3994. doi:10.1093/bioinformatics/btab634
20. Dao D, Fraser AN, Hung J, Ljosa V, Singh S, Carpenter AE. CellProfiler analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics.* 2016;32(20):3210–3212. doi:10.1093/bioinformatics/btw390
21. Fan L, Wang R, Chi C, et al. Systemic immune-inflammation index predicts the combined clinical outcome after sequential therapy with abiraterone and docetaxel for metastatic castration-resistant prostate cancer patients. *Prostate.* 2018;78(4):250–256. doi:10.1002/pros.23465
22. Dermine S, Leconte M, Leblanc S, et al. Outcomes of esophagectomy after noncurative endoscopic resection of early esophageal cancer. *Ther Adv Gastroenterol.* 2019;12:1756284819892556. doi:10.1177/1756284819892556
23. An W, Pang TT, Sun C, et al. Outcomes of endoscopic submucosal dissection versus esophagectomy for poorly differentiated superficial esophageal squamous cell carcinoma: a 10-year cohort study. *Digest Liver Dis.* 2024. doi:10.1016/j.dld.2024.06.002
24. Abe S, Hirai Y, Uozumi T, et al. Endoscopic resection of esophageal squamous cell carcinoma: current indications and treatment outcomes. *DEN Open.* 2022;2(1):e45. doi:10.1002/deo.2.45
25. Hölscher DL, Bouteldja N, Joodaki M, et al. Next-generation morphometry for pathomics-data mining in histopathology. *Nat Commun.* 2023;14(1):470. doi:10.1038/s41467-023-36173-0
26. Brown CS, Gwilliam N, Kyriillos A, et al. Predictors of pathologic upstaging in early esophageal adenocarcinoma: results from the national cancer database. *Am J Surg.* 2018;216(1):124–130. doi:10.1016/j.amjsurg.2017.07.015
27. Sgourakis G, Gockel I, Lang H. Endoscopic and surgical resection of T1a/T1b esophageal neoplasms: a systematic review. *World J Gastroenterol.* 2013;19(9):1424–1437. doi:10.3748/wjg.v19.i9.1424
28. Wu J, Hu Y, Xu L. Positive circumferential resection margin in locally advanced esophageal cancer: an updated systematic review and meta-analysis. *Updates Surg.* 2022;74(4):1187–1197. doi:10.1007/s13304-022-01256-y
29. Schlick CJR, Khorfan R, Odell DD, Merkow RP, Bentrem DJ. Margin positivity in resectable esophageal cancer: are there modifiable risk factors? *Ann Surg Oncol.* 2020;27(5):1496–1507. doi:10.1245/s10434-019-08176-z
30. Bissolati M, Desio M, Rosa F, et al. Risk factor analysis for involvement of resection margins in gastric and esophagogastric junction cancer: an Italian multicenter study. *Gastric Cancer.* 2017;20(1):70–82. doi:10.1007/s10120-015-0589-6
31. Dhingra S, Bahdi F, May SB, Othman MO. Clinicopathologic correlations of superficial esophageal adenocarcinoma in endoscopic submucosal dissection specimens. *Diagn Pathol.* 2021;16(1):111. doi:10.1186/s13000-021-01169-1
32. Uhlenhopp DJ, Then EO, Sunkara T, Gaduputi V. Epidemiology of esophageal cancer: update in global trends, etiology and risk factors. *Clin j gastroenterol.* 2020;13(6):1010–1021. doi:10.1007/s12328-020-01237-x

33. Mansour NM, Groth SS, Anandasabapathy S. Esophageal adenocarcinoma: screening, surveillance, and management. *Annual Rev Med.* 2017;68:213–227. doi:10.1146/annurev-med-050715-104218
34. Gockel I, Sgourakis G, Lyros O, Hansen T, Lang H. Dissection of lymph node metastases in esophageal cancer. *Expert Rev Anticancer Ther.* 2011;11(4):571–578. doi:10.1586/era.10.157
35. Law S, Wong J. Lymph node dissection in surgical treatment of esophageal neoplasms. *Surg Oncol Clin North America.* 2007;16(1):115–131. doi:10.1016/j.soc.2006.10.008

International Journal of General Medicine

Dovepress

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>