A Radiomics Nomogram for Non-Invasive Prediction of Progression-Free Survival in Esophageal Squamous Cell Carcinoma

Ting Yan
Shanxi Medical University
lili liu
Shanxi Medical University
Meilan Peng
Shanxi Medical University
Zhenpeng Yan
Shanxi Medical University
Qingyu Wang
Taiyuan University of Technology
Shan Zhang
Taiyuan University of Technology
Lu Wang
Shanxi Medical University
Xiaofei Zhuang
Shanxi Cancer Hospital
Huijuan Liu
Shanxi Medical University
Yanchun Ma
Shanxi Medical University
Bin Wang (wangbin01@tyut.edu.cn)
Taiyuan University of Technology
Yongping Cui
Shanxi Medical University

Research

Keywords: Esophageal squamous cell carcinoma, Computed tomography, progression-free survival, Radiomics, Nomogram

Posted Date: August 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-557179/v2

Abstract

Objectives: To construct a prognostic model for preoperative prediction based on computed tomography (CT) images of esophageal squamous cell carcinoma (ESCC).

Methods: Radiomics signature was constructed using the least absolute shrinkage and selection operator (LASSO) with high throughput radiomics features that extracted from the CT images of 272 patients (204 in training and 68 in validation cohort), who were pathologically confirmed ESCC. Multivariable logistic regression was adopted to build the radiomics signature and another predictive nomogram model, which was composed with radiomics signature, traditional TNM stage and clinical features. Then its performance was assessed by the calibration and decision curve analysis (DCA).

Results: 16 radiomics features were selected from 954 to build a radiomics signature which were significantly associated with progression-free survival (PFS) (p<0.001). The area under the curve (AUC) of performance was 0.891 (95% CI: 0.845-0.936) for training cohort and 0.706 (95% CI: 0.583-0.829) for validation cohort. The radscore of signatures' combination showed significant discrimination for survival status in both two cohort. Kaplan-Meier survival curve further confirmed the radscore has a better prognostic performance in training cohort. Radiomics nomogram combined radscore with TNM staging showed significant improvement over TNM staging alone in training cohort (C-index, 0.802 *vs* 0.628; p<0.05), and it is the same with clinical data (C-index, 0.798 *vs* 0.660; p<0.05). Findings were confirmed in the validation cohort. DCA showed CT-based radiomics model will receive benefit when the threshold probability was between 0 and 0.9. Heat maps revealed associations between radiomics features and tumor stages.

Conclusions: Multiparametric CT-based radiomics nomograms provided improved prognostic ability in ESCC.

Introduction

Esophageal cancer (EC) remains the seventh most frequent cancer and the sixth most prevalent cause of cancer deaths globally [1]. It had an estimated 477,900 new cases and 375,000 annual deaths occurring in China, and most of them are esophageal squamous cell carcinoma (ESCC) [2]. The majority of ESCC patients are diagnosed as advanced stage due to unclear early symptoms, and the 5-year survival rate is very low (less than 20%) [3, 4]. Although surgery is still the most sanative treatment, the 5-year survival rate of resectable EC treated with surgery alone is only 34–36% [5]. Hence, effective means to preoperatively predict the prognosis for ESCC patients is necessary.

Prognosis survival evaluation of EC mainly depends on traditional TNM staging for the moment. However, the TNM system only considers anatomical features and neglects the intrinsic factors of tumor, resulting in the inaccurate prognosis [6]. Then scholars started to collect clinical data as well, such as age, gender, body mass index (BMI), and quality of life [7–10]. However, the performance is still weak, for they failed to reflect the internals of tumors. Furthermore, prognostic evaluation by multi-omics

approaches is based on molecular features of a small portion of tumor tissue, which limited the understanding of the heterogeneous tumor.

Radiomics, a noninvasive, quantitative, and low-cost approach, can objectively and comprehensively evaluate tumor heterogeneity by converting medical images into high-dimensional, mineable, and quantitative imaging features via high-throughput extraction of data-characterization algorithms [11, 12]. These features have potential to reveal disease progression, thereby provide valuable information for personalized therapy and decision-support [13–23]. Previous studies have shown radiomics signature alone or merged with clinical parameters could enhance predictive accuracy in cancers [24–26]. Recently, the most widely-used imaging modality in radiomics is computed tomography (CT), which is universally used for preoperative diagnostics of ESCC. Due to the poor contrast resolution, it is difficult to distinguish the different histologic layers of esophageal wall. However, it is believed that there is still a lot of digital information can be deeply excavated through radiomics approaches.

In this study, we developed CT-based radiomics as a novel approach for individualized, pretreatment evaluation of progression-free survival (PFS) in ESCC patients (stage I-III). Additionally, we sought to reveal the association between radiomics and clinical informations.

Materials And Methods

Patients and clinical characteristics

Shanxi Medical University Review Board approved this retrospective study. The entire cohort of this study was acquired from February 2016 to October 2018 records of the Institutional Picture Archiving and Communication System (PACS), which was used to identify the patients who had histologically confirmed ESCC (TNM stage: I-III) and underwent surgery after diagnosis at Shanxi Cancer Hospital. All patients underwent pretreatment CT scans from neck to abdomen, and signed their own informed consent. All methods were carried out in accordance with relevant guidelines and regulations.

To determine the patients that could be included, we developed the following criteria: 1) pathologically confirmed ESCC; 2) underwent surgery for ESCC; 3) standard contrast-enhanced CT was performed preoperatively; and 4) complete clinical and follow-up information was available. We randomly divided the patients into training and validation cohort by a ratio of about 3:1. We trained models in training cohort and validated them in validation cohort.

Clinical characteristics including age, gender, tumor location (upper, middle, lower), drinking history, smoking history, genetic alterations, and pathologic characteristics including depth of invasion, TNM stage and lymph node metastasis informations were collected from patient records. These clinicopathologic characteristics are presented in Table 1.

Table 1
Patient and tumor characteristics in the training and validation cohorts

	Train（N=204）	Validation（N=68）
Gender		
Male	140（68.6%）	52（76.5%）
Female	64（31.4%）	17（23.5%）
Age		
Median(interquartile range)	60.5	59.5
≤56	57（27.9%）	26（36.8%）
56-66	94（46.1%）	25（36.8%）
≥66	53（26.0%）	18（26.4%）
Location		
Up	12（5.9%）	3（4.4%）
Mid	142（69.6%）	41（60.3%）
Down	50（24.5%）	24（35.3%）
Drinking		
Yes	73（35.8%）	25（36.8%）
No	131（64.2%）	44（63.2%）
Smoking		
Yes	118（57.8%）	42（61.8%）
No	86（42.2%）	27（38.2%）
Genetic History		
Yes	66（32.4%）	20（29.4%）
No	138（67.6%）	49（70.6%）
Invasion Degree		
Full layer	117（57.4%）	45（66.2%）
Non-full layer	87（42.6%）	24（33.8%）
TNM		
I	24（11.8%）	4（6%）
II	109（53.4%）	32（47.0%）

III	71（34.8%）	32（47.0%）
Lymph Node Metastasis		
Yes	80（39.2%）	32（47.1%）
No	124（60.8%）	37（52.9%）

Follow up and clinical endpoint

All patients were followed up every 1–3 months during the first 2 years, every 6 months in year 2–5, and annually thereafter. To provide an efficient tool, which would allow earlier personalized treatment, we chose PFS as the endpoint [27]. We defined PFS from the first day of treatment to the date of disease progression (locoregional recurrences or distant metastases), death from any cause, or the date of the last follow-up visit (censored). The minimum follow-up time to ascertain the PFS was 6 months.

CT acquisition and segmentation

All patients were performed the contrast-enhanced CT by using a 64-channel multi-detector CT scanner (LightSpeed VCT, GE Medical Systems, Milwaukee, Wis, USA). The acquisition parameters were as follows: 120 kV; 160 mA; 0.5-second rotation time; detector collimation: 64×0.625 mm; field of view: 350 mm×350 mm; and matrix: 512×512. After routine non-enhanced CT, contrast-enhanced CT was performed after a 25-second delay following intravenous administration of 85 mL of iodinated contrast material (Ultravist 370; Bayer Schering Pharma, Berlin, Germany) at a rate of 3.0 mL/s with a pump injector (Ulrich CT Plus 150, Ulrich Medical, Ulm, Germany). All images were reconstructed with a thick slice of 5.0 mm. For feature selection, we converted image format from DICOM to NII without applying any preprocessing.

Note that segmentation is required before the extraction of quantitative radiomics features, we performed three-dimensional manual segmentation by using 3D-Slicer software (https://www.slicer.org/), which is an open platform for medical image processing. The chief physician of Shanxi Cancer Hospital with more than five years' experience in interpreting chest radiology outlined the tumor regions for each CT image layer, and the tumor segmentation was guided and verified by the specialist. The region of interest (ROI) covered the whole tumor mass and was delineated on each CT slice, and would be used in subsequent feature extraction.

Selection of radiomics feature and building of radiomics signature

We performed the calculation through our homemade Python scripts (Python3.6, https://www.python.org) for radiomics feature extraction based on the segmentation results. A total of 954 features were obtained by calling feature calculation in pyradiomics package (open-source python package; https://pyradiomics.readthedocs.io/en/latest/), which included the following 4 categories: 1)

first-order statistics features; 2) size- and shape-based features; 3) texture features; and 4) wavelet features; and 5 typical matrixes: Gray-Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Gray Level Dependence Matrix (GLDM) and Neigbouring Gray Tone Difference Matrix (NGTDM).

We built the radiomics signature with selected features in training cohort. To reduce over-fitting or any types of bias, we applied following 2 steps: First, the best features based on univariate statistical tests (2-sample *t*-test) between death and censoring groups in the primary cohort were selected and executed by using Matlab 2016b (Mathworks, Natick, USA). Second, we used our homemade R scripts to select features that were most significant by using the least absolute shrinkage and selection operator (LASSO) method, which would be a suitable methodology for the feature selection through regression of high-demensional data (R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org, 2016). The glmnet R-packages was applied for logistic regression (open-source R package; https://cran.r-project.org/web/packages/glmnet/index.html). Additionally, the accuracy of prediction model could be improved by regularizing the features through penalized estimation. We added the L1 penalty term to the normal linear model and the parameter lambda controls the complexity of regression. When the λ was large, it indicated that there was no effect on the estimated regression parameters; while as the λ getted smaller, most covariate coefficients were shrunk to zero. Then the remaining variables with non zero coefficients were selected by the λ that the 10-fold cross-validation error was the smallest [28] [29].

Finally, the radiomics signature was built by combining those variables in the primary cohort and validated in the validation cohort. The radiomics signature is a linear combination of selected features with respective weights, which would be calculated as a factor (Radiomics score, Rad-score) for the further prediction model. The assessment method of the logistic regression model is the receiver operating characteristic (ROC) curve and its area under the curve (AUC).

Prognostic validation of radiomics signature

We calculated Rad-score for each ESCC patient and grouped them according to the following 2 rules. 1) The patients were divided into high-risk and low-risk groups based on the median Rad-score. 2) Patients with median scores were placed in high-risk groups. The radiomics signature discriminative performance of the survival status was assessed according to the overall distribution of ESCC patients. And then, the potential association of radiomics signature and clinical feature with PFS was assessed in the training cohort and validated in the validation cohort. Kaplan-Meier survival analysis was used in these two cohorts (the survival R-package was used for Kaplan-Meier survival analyses; https://cran.r-project.org/web/packages/survival/index.html). Stratified analyses were implemented to determine the PFS in subgroups of high-risk and low-risk patients. Univariate Cox Proportional Hazards Models were performed to explore the C-index of the radiomics signature (the rms R-package was used for Cox proportional hazards regression; https://cran.r-project.org/web/packages/rms/).

Performance of TNM staging and clinical nomograms in the training cohort before and after addition of Rad-score

The nomogram with the predicting model was based on the multivariable logistic regression analysis. The following candidate factors: TNM stage (dummy variable: “0” for I, “1” for II, “2” for III), status of clinical features and Rad-scores were involved in a diagnostic model for preoperative prediction of ESCC. The nomogram is a graphical representation of this prediction model in the training cohort. The prognostic performance of TNM staging and clinical nomograms in the training cohort before and after the addition of Rad-score was quantitatively measured by using harrell’s concordance index (C-Index), which is commonly used to evaluate the discriminative power of prognostic models [30]. The value of the C-index could range from 0.5, which indicated no discriminative ability, to 1.0, which indicated perfect ability to distinguish between the patients who sufferred disease progression or death and those who did not. Bootstrap analyses with 1,000 resamples were used to obtain a C-index with 95% confidence interval (CI) [31] that were corrected for potential overfitting. The calibration curves were drawn for assessing the agreement between the predicted probability of 3-year PFS and actual 3-year PFS [32].

Nomogram validation in validation cohort

The prognostic performance of TNM staging and clinical nomograms in the validation cohort before and after the addition of Rad-score was tested by the above method. Calibration curve and C-index were calculated through multivariable Cox proportional hazard regression analyses. The decision curve analysis (DCA) was introduced to evaluate the quantified net benefit of our prediction model in the validation cohort [33, 34].

Association of radiomics features with clinical data

A heat map analysis was used to evaluate the associations between clinical data and radiomics features (the gplots and pheatmap packages were used for heat maps).

Results

Clinical characteristics of all the patients

A total of 272 consecutive patients met the criteria (192 men and 80 women; mean age, 60.25 years ± 7.43) were included and divided into two cohorts by a ratio of 3:1 using computer-generated random numbers. 204 patients were enrolled in the training cohort (140 men and 64 women; mean age, 60.47 years ± 7.25), while 68 patients were enrolled in the independent validation cohort (52 men and 16 women; mean age, 59.58 years ± 7.96). The clinical characteristics with statistics of the training and validation cohorts are summarized in Table 1. No significant differences were found between these two cohorts in terms of gender, age, history of smoking and drinking, location, genetic history□invasion degree□lymph node metastasis and overall TNM Stage (p = 0.152-0.904). The median PFS was 36 months (range, 6-75 months).

Radiomics feature selection and radiomics signature building

A total of 954 features were extracted from CT images, and might contain many redundant and highly correlated features. To find out robust and valuable features, the following steps were performed: Firstly, 211 features were selected by univariate statistical tests ($p<0.05$) (Table 2). Then, based on the LASSO logistic regression algorithm approach in the training cohort, we selected the features with non-zero coefficients. As a result, 16 radiomics features were screened out from 211 features (Table 3). The procedures of parameter tuning and feature space reduction of the regression model are illustrated in Fig. 1. To build the radiomics signature, the 16 features were selected and involved in the Rad-score-based prognostic model. The discriminative ability of the survival status based on radiomics signatures was assessed by ROC in the both cohorts respectively (Fig. 2a).

Table 2

Radiomics features selection results based on the Anova

Result category	CT
Number of selected features	221
The best-performance feature	HLL-original_glcm_InverseVariance (P=2.316589e-04)

Table 3

Radiomics signature selection results with descriptions

Future name	Future coefficient
HHH_ngtdm_Busyness	0.00
HHL_firstorder_Skewness	0.09
HLH_firstorder_Median	-1.53
HLH_glszm_SmallAreaEmphasis	-7.75
HLL_glcm_ClusterShade	0.00
HLL_glcm_InverseVariance	-9.92
HLL_glszm_SizeZoneNonUniformityNormalized	8.24
LHH_gldm_DependenceNonUniformityNormalized	31.68
LHH_ngtdm_Busyness	0.00
LHL_glcm_Idn	17.74
LHL_glszm_SmallAreaLowGrayLevelEmphasis	-2.42
LHL_gldm_SmallDependenceLowGrayLevelEmphasis	-134.61
LLH_glcm_Contrast	0.04
LLL_glszm_LargeAreaHighGrayLevelEmphasis	0.00
ORI_glszm_LowGrayLevelZoneEmphasis	-2.16
ORI_gldm_LargeDependenceLowGrayLevelEmphasis	-0.09

1) Median：The median gray level intensity within ROI.

2) Skewness：The asymmetric distribution of the Mean value. Depending on where the tail is elongated and the mass of distribution is concentrated, it can be positive or negative.

3) Cluster Shade: A measure of skewness and uniformity of the GLCM. A higher cluster shade implies greater asymmetry about the mean.

4) IDN (inverse difference normalized): Another measure of local homogeneity of images. Unlike Homogeneity1, IDN normalizes the difference between neighboring intensity values by dividing over the total number of discrete intensity values.

5) Contrast: A measure of local intensity variation, favoring values away from the diagonal ($i=j$). A larger value correlates with a greater disparity in intensity values among neighboring voxels.

6) Small Area Emphasis (SAE): A measure of the distribution of small size zones, with a greater value indicative of more smaller size zones and more fine textures.

7) SizeZoneNonUniformityNormalized: The variability of size zone volumes throughout images, with a lower value indicating more homogeneity among zone size volumes in images. it's the normalized version of the SZN formula.

8) Small Dependence Low Gray Level Emphasis: The proportion in images of the joint distribution of smaller size zones with lower gray-level values.

9) LAHGLE: The proportion in images of the joint distribution of larger size zones with higher gray-level values.

10) LGLZE: Distribution of lower gray-level size zones, a higher value indicating a greater proportion of lower gray-level values and size zones in images.

11) Dependence Non-Uniformity Normalized (DNN): Measures the similarity of dependence throughout images, with a lower value indicating more homogeneity among dependencies in images. This is the normalized version of the DLN formula.

12) Small Dependence Low Gray Level Emphasis (SDLGLE): Measures the joint distribution of small dependence with lower gray-level values.

13) Large Dependence Low Gray Level Emphasis (LDLGLE): Measures the joint distribution of large dependence with lower gray-level values.

14) Busyness: A measure of the change from a pixel to its neighbour. A high value for busyness indicates a 'busy' image, with rapid changes of intensity between pixels and its neighbourhood.

Prognostic validation of radiomics signature

Rad-score for each patient in the training cohort and validation cohort correspondingly showed that the higher the Rad-score, the greater the probability of death (Fig. 2b, c). Besides, in the training cohort, the radiomics signature from CT images yielded the highest C-index, which was 0.785 (95% CI: 0.719 to 0.850). In the validation cohort, the radiomics signature from CT images yielded a C-index of 0.692 (95% CI: 0.589 to 0.794). It showed a significant discrimination between the PFS of high-risk and low-risk patients in subgroup analyses (Fig. 3).

Performance of TNM staging and clinical nomograms in the training cohort before and after the addition of Rad-score

We developed a radiomics nomogram that integrated the radiomics signature from the CT images with the traditional TNM staging system, which yielded a C-index of 0.628 (95% CI: 0.570 to 0.687). This nomogram significantly improved the discrimination ability in evaluating PFS (C-index: 0.802; 95% CI: 0.737 to 0.868) than TNM staging system ($p < 0.05$; Fig. 4a), and showed good calibration as well (Fig. 4b). Moreover, a radiomics nomogram was created by integrating the radiomics signature from the CT images with all clinical data, whose nomogram yielded a C-index of 0.660 (95% CI: 0.595 to 0.726). We found that the radiomics nomogram possessed good calibration and seemed to be more accurate than clinical nomogram for evaluating PFS (C-index: 0.799; 95% CI: 0.733 to 0.864) with a p-value < 0.05 (Fig. 4c, d).

The validation of nomograms in validation cohort

In validation cohort, the C-index of traditional TNM staging system is 0.516 (95% CI: 0.424 to 0.607). We integrated the radiomics signature with the TNM staging system to produce a radiomics nomogram. It showed an improvement over the TNM staging system alone (C-index: 0.691; 95% CI: 0.588 to 0.794). The calibration curve of probability in PFS evaluation showed good agreement between nomogram-evaluated and actual observation (Figure not shown). While, the clinical nomogram yielded a C-index of 0.683 (95% CI: 0.581 to 0.786) in the validation cohort, and was advanced by combining with radiomics signature (C-index: 0.774; 95% CI: 0.671 to 0.876). The calibration curves of this nomogram showed good agreement between nomogram-evaluated and actual survival (Figure not shown). The DCA for the prediction model derived from the addition of Rad-score before and after is presented in Fig. 5a. It showed that the predictive model collaborated with Rad-score had a batter net benefit than that with only traditional TNM staging combined with clinical features.

Association of radiomics features with clinical data

The ESCC patients with similar patterns of radiomics expression were clustered through unsupervised clustering (Fig. 5b). Then we organized a heat map to determine the association between radiomics features and clinical data (Fig. 5b). The results showed significant correlations between signature features LHL−glcm_Idn, LHL−gldm_Small Dependence Low Gray Level Emphasis with drinking (p < 0.001) as well as gender (p < 0.03). Moreover, LHH-gldm_Dependence NonUniformity Normalized was significantly associated with gender (p < 0.001). LLH-glcm_Contrast, LHL-glcm_Idn and LHH-gldm_Dependence NonUniformity Normalized were associated with smoking (p < 0.05). LHH-gldm_Dependence NonUniformity Normalized and HHH−ngtdm_Busyness was associated with invasion degree (p = 0.04-0.05). LHH-gldm_Dependence NonUniformity Normalized and LHL−gldm_Small Dependence Low Gray Level Emphasis were associated with overall stage (p < 0.05). In contrast, no radiomics feature was significantly associated with age and metastasis (for all, p > 0.05).

Discussion

Here we firstly developed and validated a new approach basedon CT radiomics for the evaluation of PFS before treatment in ESCC (stage I-III). The radiomics signature from CT images demonstrated better prognostic performance than traditional clinical informations alone. It could be competently differentiated between patients with high-risk and low-risk, who had significantly different 3-year PFS, and were defined according to the median Rad-score. The developed radiomics nomogram transcended both the traditional TNM staging system and clinical nomogram alone.

In clinical practice, CT, magnetic resonance imaging (MRI), positron emission tomography (PET), and endoscopic ultrasound (EUS) have their own advantages and disadvantages in the staging of esophageal cancer, or even cancer. But the use of these modalities is limited for their cost in both time and money. CT own the highest cost performance for its high availability and noninvasive process. However, the traditional prognosis was depended on the doctors' observation, which is differ greatly

according to the experience. Moreover, the evaluation from traditional clinical informations is even more inadequate. It is believed that there is still a lot of digital information that can be deeply excavated through the radiomics methodology, and used for judgement conversely. Therefore, we analyzed all acquired CT images and constructed a CT-based radiomics signature. And the results confirmed our expectations that the radiomics signatures have the potential for evaluating prognosis in ESCC.

To build the radiomics signature, we selected 16 potential predictors from 954 candidate features through both selecting highly correlated features with event outcomes and LASSO logistic regression. The radiomics features obtained are generally accurate. Because the regression coefficients of most features have shrunk towards zero during model fitting. It not only allowed the identification of features that had strongest association with PFS [35], but also avoided over fitting [36]. The radiomics signatures from CT images could revealed adequate discrimination in both the training cohort (C-index, 0.785) and the validation cohort (C-index, 0.692). Additionally, the selected features were used to improve radiomics signature and Rad-scores. We sorted the Rad-scores of all the patients with the labeled living status in Fig. 2a, suggested that the Rad-score could potentially differentiate the two types of patients. Other related statistical analysis also supported that the radiomics signature could be used as a biomarker in prognosis of ESCC. Compared to the traditional TNM staging system and clinical nomogram, we found the radiomics signature took a dominating factor position in our nomogram in both the training cohort and validation cohort. It means the radiomics signature has better discrimination and prognosis ability compared to that of classical radiologists, indicating the clinical importance of our findings due to the traditional clinical information and TNM staging are routinely used in clinical practice [37, 38].

Generally, doctors are using the traditional TNM staging system for risk pridiction and treatment planning making nowadays. However, there were obvious differences in PFS with the same clinical identified disease stage, indicating that tumor heterogeneity would affect the survival outcomes. The ESCC patients (stage I-III) with shorter PFS may benefit from the prognostic model, because they may give up aggressive treatments to avoid the suffering and overspending. Here, we developed the radiomics features possessing better prognostic ability than traditional TNM staging system for pretreatment PFS in validation cohort as well as training cohort. It might because that our study was focused on ESCC patients with stage I-III tumors (Table 1), and the patients with stage I accounted for a small proportion (11.8% in training cohort, 6% in validation cohort). In consequence, it might difficult to accurately stratify PFS since the similar information of clinical stage. Moreover, the traditional TNM stage mainly reflect the clinicopathologic features of cancer patients, such as tumor size, lymph node involvement and distant metastasis status, respectively. They do have prognostic value in tumor treatment, but neglected the intratumor heterogeneity, which was deemed as a crucial factor for tumor progression and prognosis [39]. As a result, it provided an inefficient nomogram performance in both the training cohort (C-index, 0.628) and the validation cohort (C-index, 0.515). While the radiomics approach did extract the features of entire tumor from medical images, by which produce a more comprehensive way to noninvasively involve the intratumor heterogeneity. This might be why the combination of radiomics signatures and traditional TNM staging could provide a better nomogram performance in both training cohort (C-index, 0.802) and

validation cohort (C-index, 0.691). Hence, the radiomics signatures could asist prognosis for ESCC complementarily to the traditional TNM staging.

Previous studies reported that clinical infromations, including gender, pathological type, tumor differentiation, depth of invasion, and regional lymph node metastasis were associated with overall survival (OS) outcomes through univariate analysis. While multivariate analysis showed that pathologic type, depth of invasion, and regional lymph node metastasis were the independent predictors of OS [40]. Besides, the tumor volume of ESCC could be used as an important prognostic factor for radiotherapy and chemotherapy assessment [41–43]. Therefore, we exploited a clinical nomogram that combined available risk factors (age, gender, invasion degree, location, genetic history, metastasis) with overall stage, but it doesn't exibit well (C-index of training cohort, 0.683; C-index of validation cohort, 0.660). Then, we developed the nomogram by combining radiomics signature to it in both training cohort (C-index, 0.799) and validation cohort (C-index, 0.774). This process suggested that radiomics signatures have crucial prognostic value for ESCC patients.

Unlike the traditional methods, radiomics system is a noninvasive and low-spending approach, which could provide new insights into the associations between tumor intrinsic properties and biological behaviors. We analyzed the relationship between radiomics features and tumor-associated characteristics, and observed some radiomics features were related to the general information of patients (gender, drinking or smoking information, Fig. 5b). Additionally, our radiomics system showed some radiomics features were associated with invasion degree as well (Fig. 5b). As a result, the present study may provide some different insights into the mechanisms of lymphatic metastasis of ESCC, which require future investigation.

There were several limitations in our study. First, we used thick-slice CT images rather than thin-slice images for the extraction of radiomics signatures. Zhao *et al*. [44] found that thin-slice images could reflected texture features of tumor more complete than thick-slice images. For the measurement of tumor volumes, thin-slice images had less measurement variability. We will further study the effect of thin-slice CT images for the staging of ESCC and confirm whether the performance is comparable with thick-slice images. Second, all data involved in this study are derived from the same hospital, resulting the lack of multi-center validation. The further investigations on the applicability to the patients of other institutions is still required. Third,the analysis did not cover two-way or higher-order interactions of the radiomics features. If the interaction(s) strongly associated with the outcomes were applied, the prognostic performance of our nomogram may be significantly improved. However, to reveal the interactions of multiple factors is challenging. In brief, our study clearly showed that the radiomics approach is potential for the prognosis of ESCC patients.

Conclusions

In this study, the radiomics signature from CT images showed great prognostic performance in progression-free survival ($p < 0.001$) for ESCC. Our results showed that the radiomics nomogram

combined radscore with TNM staging significantly improved over TNM staging alone. Therefore, the multiparametric CT-based radiomics nomograms provided improved prognostic ability in ESCC.

Abbreviations

CT: Computed tomography; EC: Esophageal cancer; ESCC: Esophageal squamous cell carcinoma; LASSO: Least absolute shrinkage and selection operator; DCA: Decision curve analysis; PFS: Progression-free survival; AUC: Area under the curve; BMI: Body mass index; PACS: Picture archiving and communication system; ROI: Region of interest□ROC: Receiver operating characteristic; GLCM: Gray-Level Co-occurrence Matrix; GLRLM: Gray Level Run Length Matrix; GLSZM: Gray Level Size Zone Matrix; GLDM: Gray Level Dependence Matrix; NGTDM: Neigbouring Gray Tone Difference Matrix; C-Index: Concordance index; CI: Confidence interval; DCA: Decision curve analysis; MRI: Magnetic resonance imaging; PET: Positron emission tomography; EUS: Endoscopic ultrasound; OS: Overall survival.

Declarations

Funding

This work was supported by funding from the National Natural Science Foundation of China (81702449), the Foundation for Youths of Shanxi Province (201801D221400), the Fund for Shanxi "1331 Project" and "1331 Project" Key Subjects Construction.

Author Contributions

T.Y. conceived the study, designed the experiments, analyzed the data and wrote the manuscript. B.W. and Y.P.C. edited the manuscript. L.L.L., M.L.P., Z.P.Y., Q.Y.W. and S.Z. supervised data analysis. X.F.Z. provided clinical informations, coordinated and performed segmentation of CT images. L.W., H.J.L. and Y.C.M. performed the statistics analyses. All authors had access to the study data and reviewed and approved the final manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Data access and integrity

The authors declare that they had full access to all the data in this study and the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Ethics approval and consent to participate

The dataset of this study was approved by the ethics board of Shanxi Medical University (2013101)，and all the particitatants signed the consent.

Consent for publication

Not applicable.

Competing Interests

All of the authors declare no personal, professional, and financial conflicts of interest.

Author details

[1]Department of Pathology & Shanxi Key Laboratory of Carcinogenesis and Translational Research on Esophageal Cancer, Shanxi Medical University, Taiyuan 030001, P.R. China.[2]College of Information and Computer, Taiyuan University of Technology, Taiyuan 030001, P.R. China.[3]Department of Thoracic Surgery, Shanxi Cancer Hospital, 030013 Taiyuan, P.R. China.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424. doi:https://doi.org/10.3322/caac.21492.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66(2):115–32. doi:https://doi.org/10.3322/caac.21338.
3. Chen W, Zeng H, Chen R, Xia R, Yang Z, Xia C, Zheng R, Wei W, Zhuang G, Yu X, et al. Evaluating efficacy of screening for upper gastrointestinal cancer in China: a study protocol for a randomized controlled trial. Chin J Cancer Res. 2017;29(4):294–302. doi:https://doi.org/10.21147/j.issn.1000-9604.2017.04.02.
4. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma *Lancet*. 2013;381(9864):400–12. doi:https://doi.org/10.1016/S0140-6736(12)60643-6.
5. Omloo JM, Lagarde SM, Hulscher JB, Reitsma JB, Fockens P, van Dekken H, Ten Kate FJ, Obertop H, Tilanus HW, van Lanschot JJ. Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the mid/distal esophagus: five-year survival of a randomized clinical trial. *Ann Surg*.2007; 246(6):992–1000; discussion 1000-1. doi: https://doi.org/10.1097/SLA.0b013e31815c4037.

6. Wang HY, Sun BY, Zhu ZH, Chang ET, To KF, Hwang JS, Jiang H, Kam MK, Chen G, Cheah SL, et al. Eight-signature classifier for prediction of nasopharyngeal [corrected] carcinoma survival. J Clin Oncol. 2011;29(34):4516–25. doi:https://doi.org/10.1200/JCO.2010.33.7741.
7. Ng SH, Lin CY, Chan SC, Lin YC, Yen TC, Liao CT, Chang JT, Ko SF, Wang HM, Chang CJ, et al. Clinical utility of multimodality imaging with dynamic contrast-enhanced MRI, diffusion-weighted MRI, and 18F-FDG PET/CT for the prediction of neck control in oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiation. PLoS One. 2014;9(12):e115933. doi:https://doi.org/10.1371/journal.pone.0115933.
8. Tang LQ, Chen QY, Fan W, Liu H, Zhang L, Guo L, Luo DH, Huang PY, Zhang X, Lin XP, et al. Prospective study of tailoring whole-body dual-modality [18F]fluorodeoxyglucose positron emission tomography/computed tomography with plasma Epstein-Barr virus DNA for detecting distant metastasis in endemic nasopharyngeal carcinoma at initial staging. J Clin Oncol. 2013;31(23):2861–9. doi:https://doi.org/10.1200/jco.2012.46.0816.
9. Zeng L, Guo P, Li JG, Han F, Li Q, Lu Y, Deng XW, Zhang QY, Lu TX. Prognostic score models for survival of nasopharyngeal carcinoma patients treated with intensity-modulated radiotherapy and chemotherapy. Oncotarget. 2015;6(36):39373–83. doi:https://doi.org/10.18632/oncotarget.5781.
10. Zhang M, Wei S, Su L, Lv W, Hong J. Prognostic significance of pretreated serum lactate dehydrogenase level in nasopharyngeal carcinoma among Chinese population: A meta-analysis. Med (Baltim). 2016;95(35):e4494. doi:https://doi.org/10.1097/md.0000000000004494.
11. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun.2014; 54006. doi:https://doi.org/10.1038/ncomms5006.
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures. They Are Data *Radiology*. 2016;278(2):563–77. doi:https://doi.org/10.1148/radiol.2015151169.
13. Cameron A, Khalvati F, Haider MA, Wong A. MAPS: A Quantitative Radiomics Approach for Prostate Cancer Detection. IEEE Trans Biomed Eng. 2016;63(6):1145–56. doi:https://doi.org/10.1109/TBME.2015.2485779.
14. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging. 2013;40(1):133–40. doi:https://doi.org/10.1007/s00259-012-2247-0.
15. Ginsburg SB, Algohary A, Pahwa S, Gulani V, Ponsky L, Aronen HJ, Boström PJ, Böhm M, Haynes AM, Brenner P, et al. Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study. J Magn Reson Imaging. 2017;46(1):184–93. doi:https://doi.org/10.1002/jmri.25562.
16. Huynh E, Coroller TP, Narayan V, Agrawal V, Hou Y, Romano J, Franco I, Mak RH, Aerts HJ. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. Radiother Oncol. 2016;120(2):258–66. doi:https://doi.org/10.1016/j.radonc.2016.05.024.

17. Jin JY, Kong FM. Personalized Radiation Therapy (PRT) for Lung Cancer. Adv Exp Med Biol.2016; 890175–202. doi:https://doi.org/10.1007/978-3-319-24932-2_10.
18. Kotrotsou A, Zinn PO. and R.R. Colen. Radiomics in Brain Tumors: An Emerging Technique for Characterization of Tumor Environment. Magn Reson Imaging Clin N Am. 2016;24(4):719–29. doi:https://doi.org/10.1016/j.mric.2016.06.006.
19. Lee G, Lee HY, Park H, Schiebler ML, van Beek EJR, Ohno Y, Seo JB, Leung A. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. Eur J Radiol.2017; 86297–307. doi:https://doi.org/10.1016/j.ejrad.2016.09.005.
20. Marin Z, Batchelder KA, Toner BC, Guimond L, Gerasimova-Chechkina E, Harrow AR, Arneodo A, Khalil A. Mammographic evidence of microenvironment changes in tumorous breasts. Med Phys. 2017;44(4):1324–36. doi:https://doi.org/10.1002/mp.12120.
21. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. Expert Rev Precis Med Drug Dev. 2016;1(2):207–26. doi:https://doi.org/10.1080/23808993.2016.1164013.
22. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. Br J Radiol. 2017;90(1070):20160642. doi:https://doi.org/10.1259/bjr.20160642.
23. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, Abdalah MA, Schabath MB, Goldgof DG, Mackin D, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys. 2017;44(3):1050–62. doi:https://doi.org/10.1002/mp.12123.
24. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, Liang C, Tian J, Liang C. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology. 2016;281(3):947–57. doi:https://doi.org/10.1148/radiol.2016152234.
25. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, Ma ZL, Liu ZY. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J Clin Oncol. 2016;34(18):2157–64. doi:https://doi.org/10.1200/JCO.2015.65.9128.
26. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, Lian Z, Liu J, Luo X, Pei S, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. Clin Cancer Res. 2017;23(15):4259–69. doi:https://doi.org/10.1158/1078-0432.Ccr-16-2910.
27. Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M, Labianca R, Seitz JF, O'Callaghan CJ, Francini G, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol. 2005;23(34):8664–70. doi:https://doi.org/10.1200/jco.2005.01.6071.
28. Kumamaru KK, Saboo SS, Aghayev A, Cai P, Quesada CG, George E, Hussain Z, Cai T, Rybicki FJ. CT pulmonary angiography-based scoring system to predict the prognosis of acute pulmonary embolism. J Cardiovasc Comput Tomogr. 2016;10(6):473–9. doi:https://doi.org/10.1016/j.jcct.2016.08.007.
29. Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S. Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity:

simulation and application. BMC Med Res Methodol. 2016;16(1):154. doi:https://doi.org/10.1186/s12874-016-0254-8.

30. FE HJ Hmisc: harrell miscellaneous. *R package version*.2008; 1(2).
31. B CA.a.R. boot: Bootstrap R (S-Plus) functions. *R package version*.2012; 1(7).
32. Pencina MJ, D'Agostino RB, Sr., and Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011;30(1):11–21. doi:https://doi.org/10.1002/sim.4085.
33. Shen C, Liu Z, Wang Z, Guo J, Zhang H, Wang Y, Qin J, Li H, Fang M, Tang Z, et al. Building CT Radiomics Based Nomogram for Preoperative Esophageal Cancer Patients Lymph Node Metastasis Prediction. Transl Oncol. 2018;11(3):815–24. doi:https://doi.org/10.1016/j.tranon.2018.04.005.
34. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak.2008; 853. doi:https://doi.org/10.1186/1472-6947-8-53.
35. Ndhlovu ZM, Chibnik LB, Proudfoot J, Vine S, McMullen A, Cesa K, Porichis F, Jones RB, Alvino DM, Hart MG, et al. High-dimensional immunomonitoring models of HIV-1-specific CD8 T-cell responses accurately identify subjects achieving spontaneous viral control. Blood. 2013;121(5):801–11. doi:https://doi.org/10.1182/blood-2012-06-436295.
36. Hepp T, Schmid M, Gefeller O, Waldmann E, Mayr A. Approaches to Regularized Regression - A Comparison between Gradient Boosting and the Lasso. Methods Inf Med. 2016;55(5):422–30. doi:https://doi.org/10.3414/me16-01-0033.
37. Li AC, Xiao WW, Wang L, Shen GZ, Xu AA, Cao YQ, Huang SM, Lin CG, Han F, Deng XW, et al. Risk factors and prediction-score model for distant metastasis in nasopharyngeal carcinoma treated with intensity-modulated radiotherapy. Tumour Biol. 2015;36(11):8349–57. doi:https://doi.org/10.1007/s13277-015-3574-0.
38. Wu S, Xia B, Han F, Xie R, Song T, Lu L, Yu W, Deng X, He Q, Zhao C, et al. Prognostic Nomogram for Patients with Nasopharyngeal Carcinoma after Intensity-Modulated Radiotherapy. PLoS One. 2015;10(8)::e0134491. doi:https://doi.org/10.1371/journal.pone.0134491.
39. Yan T, Cui H, Zhou Y, Yang B, Kong P, Zhang Y, Liu Y, Wang B, Cheng Y, Li J, et al. Multi-region sequencing unveils novel actionable targets and spatial heterogeneity in esophageal squamous cell carcinoma. Nat Commun. 2019;10(1):1670. doi:https://doi.org/10.1038/s41467-019-09255-1.
40. Liu Y, Chen X, Wang Y, Wang F, Gong Y, Zhang J. Clinical features and prognostic factors for surgical treatment of esophageal squamous cell carcinoma in elderly patients. J BUON. 2019;24(3):1240–4.
41. Chen CZ, Chen JZ, Li DR, Lin ZX, Zhou MZ, Li DS, Chen ZJ. Long-term outcomes and prognostic factors for patients with esophageal cancer following radiotherapy. World J Gastroenterol. 2013;19(10):1639–44. doi:https://doi.org/10.3748/wjg.v19.i10.1639.
42. Chen Y, Zhang Z, Jiang G, Zhao K. Gross tumor volume is the prognostic factor for squamous cell esophageal cancer patients treated with definitive radiotherapy. J Thorac Dis. 2016;8(6):1155–61. doi:https://doi.org/10.21037/jtd.2016.04.08.

43. Li R, Chen TW, Hu J, Guo DD, Zhang XM, Deng D, Li H, Chen XL, Tang HJ. Tumor volume of resectable adenocarcinoma of the esophagogastric junction at multidetector CT: association with regional lymph node metastasis and N stage. Radiology. 2013;269(1):130–8. doi:https://doi.org/10.1148/radiol.13122269.
44. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, Schwartz LH. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep.2016; 623428. doi:https://doi.org/10.1038/srep23428.

Figures

Figure 1

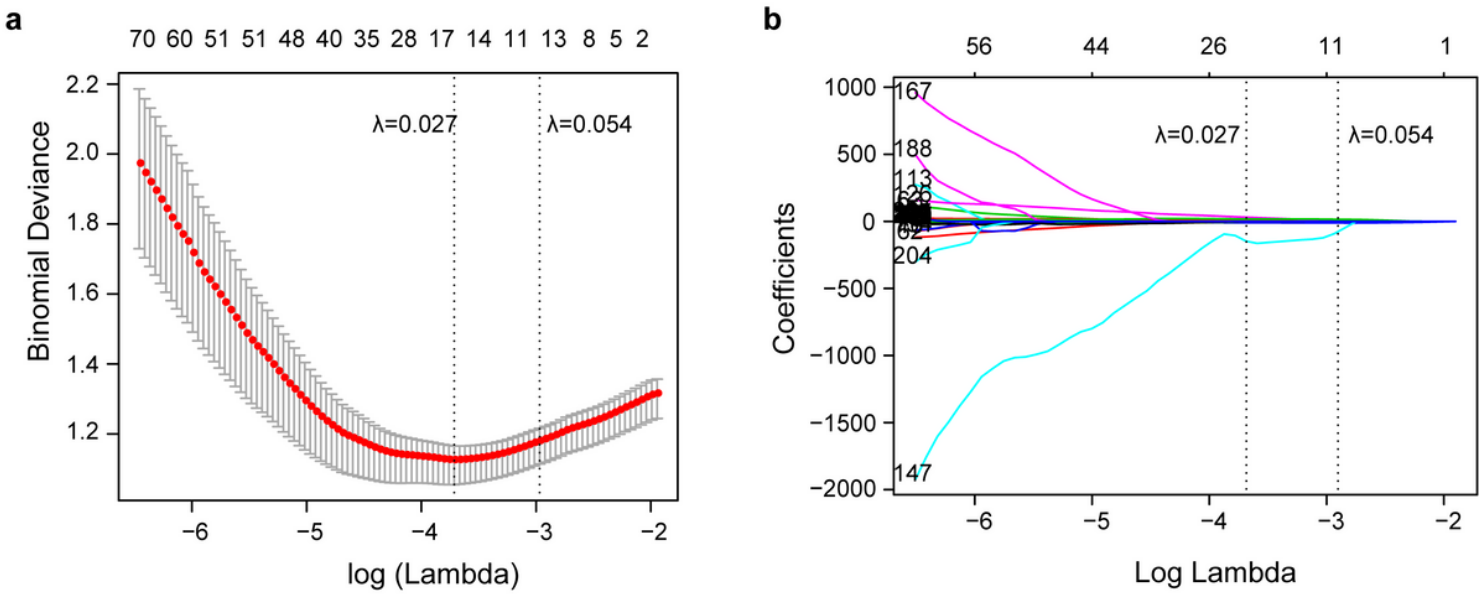

Figure 1. Radiomics feature selection using LASSO logistic regression model. (a) Identification of the optimal penalization coefficient lambda (λ) in the LASSO model used 10-fold cross-validation and the minimum criterion. As a result，a λ value of 0.027 was selected. (b) LASSO coefficient profiles of the 211 radiomics features. The dotted vertical line was plotted at the value selected using 10-fold cross-validation in Fig. 1A，for which the optimal λ resulted in eight non-zero coefficients.

Figure 1

See image above for figure legend.

Figure 2

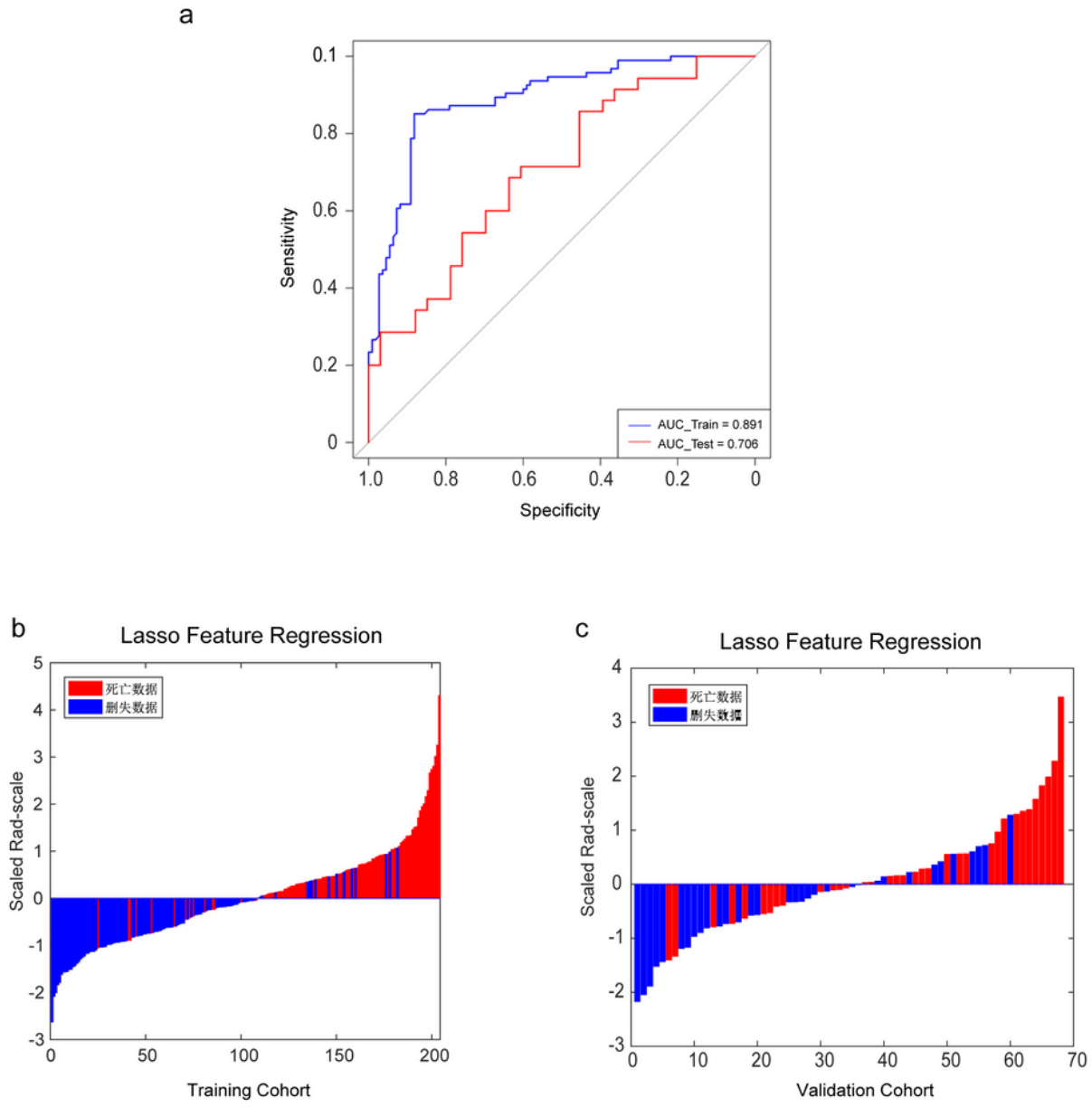

Figure 2. Rad-score for each patient in the training cohort and validation cohort. (a) ROCs were employed to assess the radiomics signature discriminative performance of the survival status. ROC in the training cohort with 0.891 (95% CI: 0.845-0.936, sensitivity = 85.1%, specificity = 88.2%); ROC in the validation cohort with 0.706 (95% CI: 0.583-0.829, sensitivity = 68.6%, specificity = 63.6%). Rad-score for each patient in the training cohort (b) and validation cohort (c). Blue bars show scores for patients who survived without disease progression or were censored, while red bars show scores for those who experienced progression or died.

Figure 2

See image above for figure legend.

Figure 3

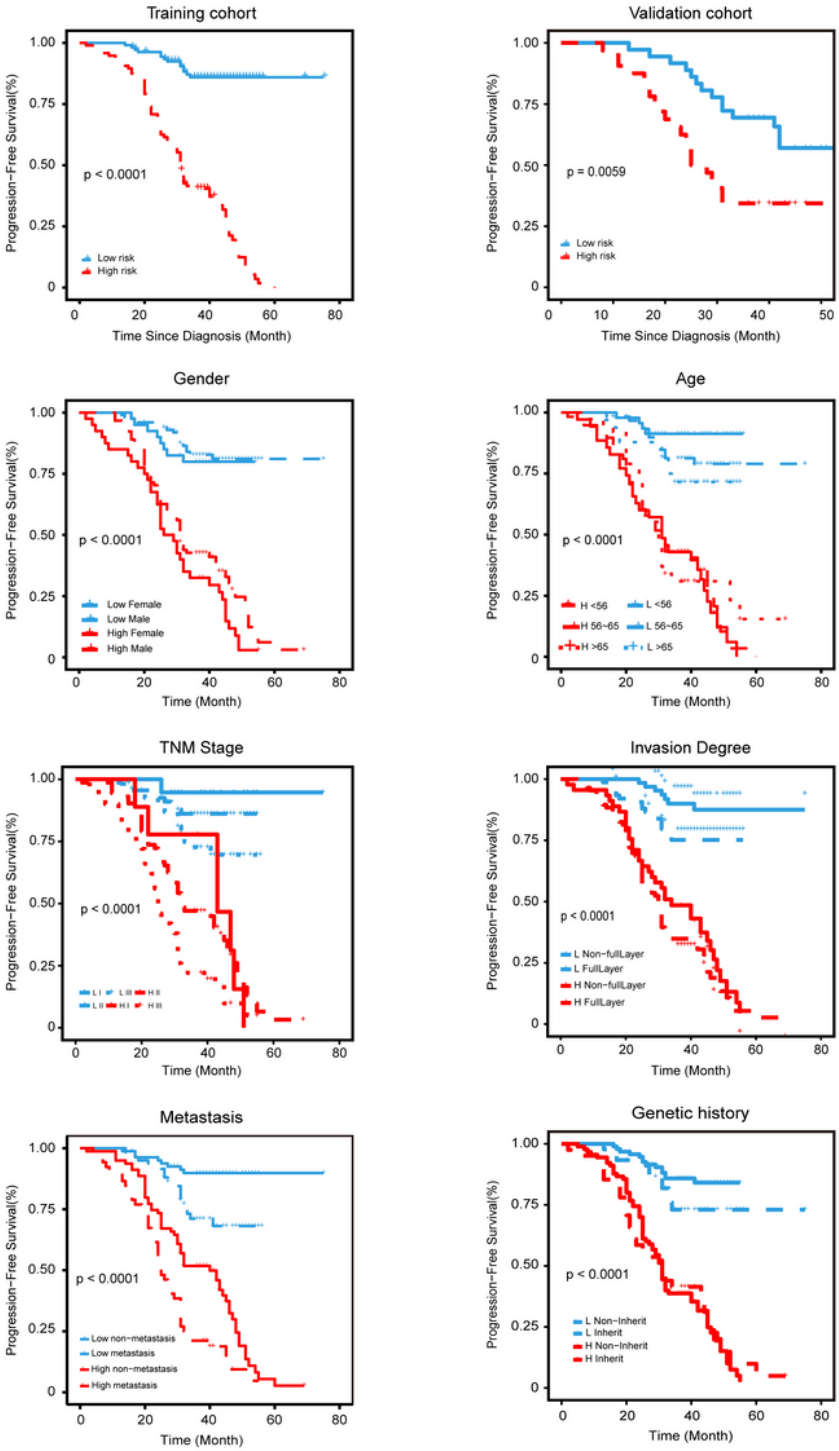

Figure 3. Stratified analyses were performed to estimate PFS in various subgroups, comparing high-risk patients and low-risk patients.

Figure 3

See image above for figure legend.

Figure 4

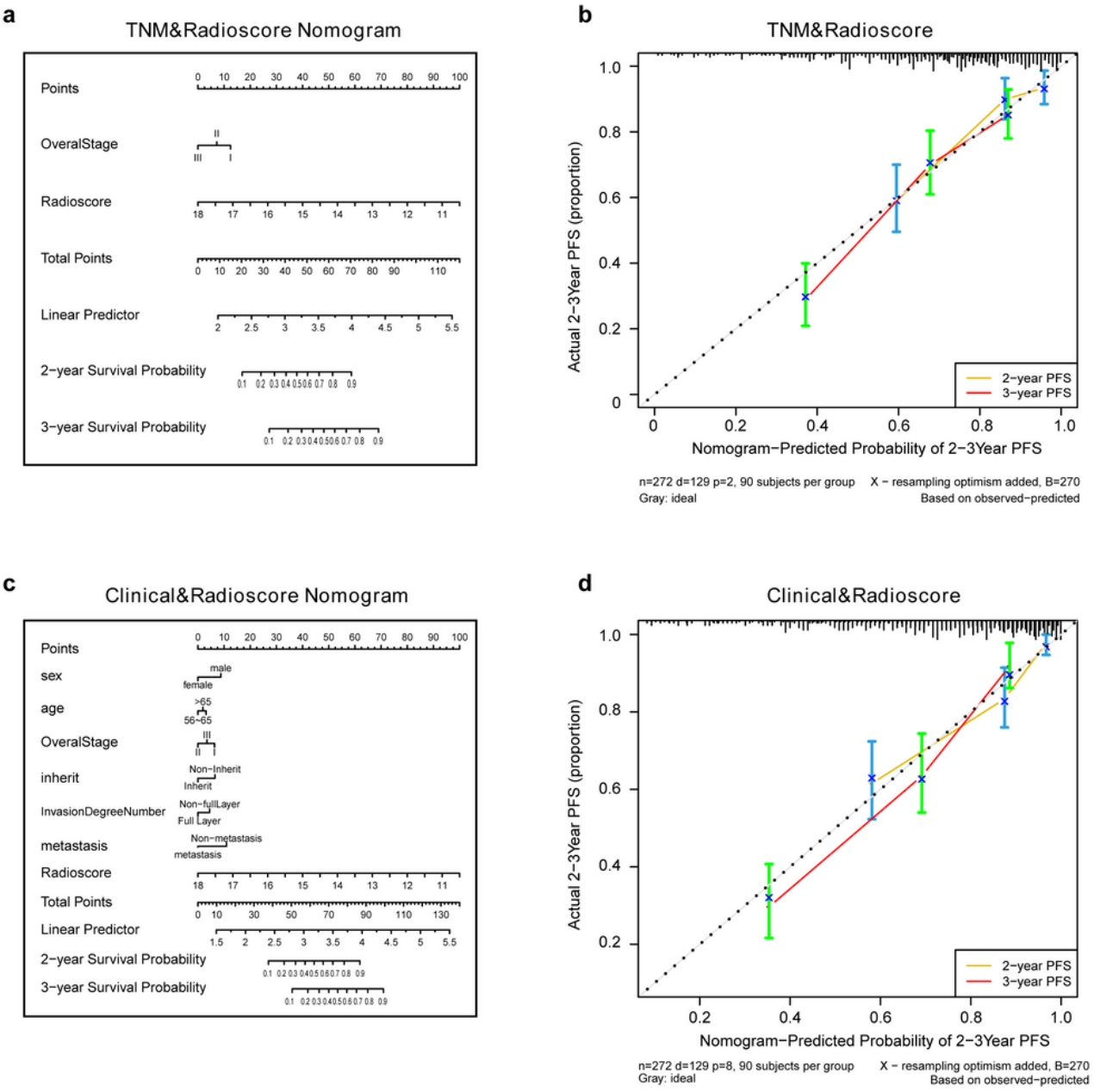

Figure 4. (a) A radiomics nomogram integrated the radiomics signature from CT images with the TNM staging system in the training cohort. (b) Calibration curve of the radiomics nomogram. The diagonal dotted line represents an ideal evaluation, while the yellow and red solid lines represent the performance of the nomogram. Closer fit to the diagonal dotted line indicates a better evaluation. (c) Adding Age, gender, invasion degree, location, genetic history, metastasis to the radiomics nomogram. (d) Calibration curve of the radiomics nomogram with the addition of Age, gender, invasion degree, location, genetic history, metastasis.

Figure 4

See image above for figure legend.

Figure 5

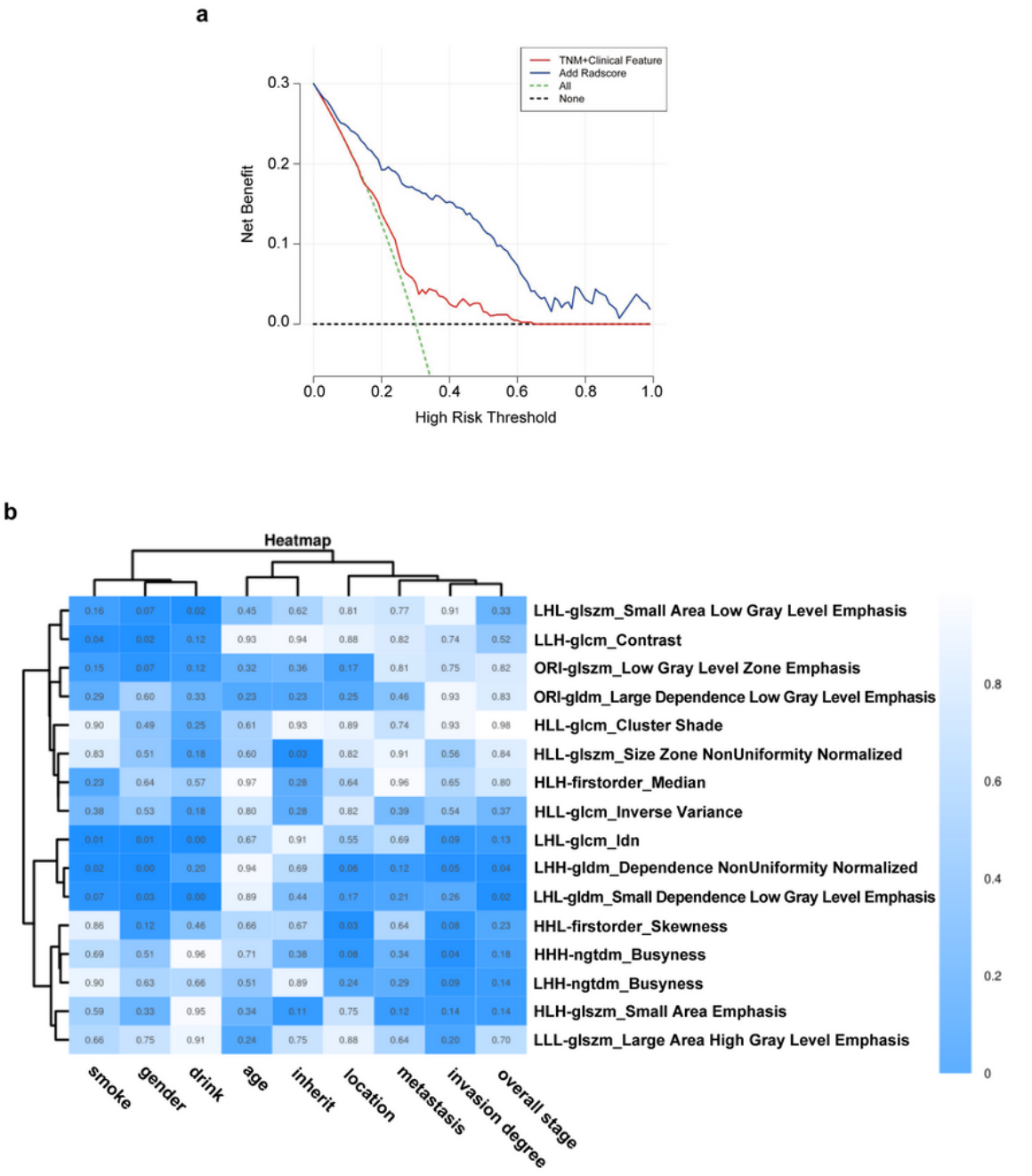

Figure 5. (a) The DCA of the radiomics-comparison-based nomogram. The black dotted line describes the scheme of no treatment. The green dotted line describes the scheme of treatment. The red line represents our predictive model with only traditional TNM staging combined with clinical features. And the blue line represents our personalized prediction model that added Rad-score. The x-axis is the threshold probability and the y-axis is the net benefit. It can be seen the personalized prediction model with Rad-score added had a batter net benefit than the traditional predictive model when the threshold is in the range of 0 to 0.9. Hence, the patient with ESCC would receive benefit from taking our CT-based radiomics nomogram guidance. (b) Heatmap of associations between selected radiomics features and clinical data. P <0.05 indicate statistically associations.as determined using t-tests.

Figure 5

See image above for figure legend.