Radiotherapy and Oncology 154 (2021) 6–13

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma

Yihuai Hu [a,b,c,1], Chenyi Xie [d,1], Hong Yang [a,b,c], Joshua W.K. Ho [e], Jing Wen [b,c], Lujun Han [b,f], Ka-On Lam [g], Ian Y.H. Wong [h], Simon Y.K. Law [h], Keith W.H. Chiu [d], Varut Vardhanabhuti [d,*], Jianhua Fu [a,b,c,*]

[a] Department of Thoracic Surgery, Sun Yat-sen University Cancer Center; [b] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine; [c] Guangdong Esophageal Cancer Institute, Guangzhou, China; [d] Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, University of Hong Kong; [e] School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China; [f] Department of Medical Imaging, Sun Yat-sen University Cancer Center, Guangzhou, China; [g] Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, University of Hong Kong; and [h] Department of Surgery, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 9 May 2020
Received in revised form 20 August 2020
Accepted 6 September 2020
Available online 15 September 2020

Keywords:
Esophageal squamous cell carcinoma
Neoadjuvant chemoradiotherapy
Deep learning
Radiomics
Computed tomography

ABSTRACT

Background: Deep learning is promising to predict treatment response. We aimed to evaluate and validate the predictive performance of the CT-based model using deep learning features for predicting pathologic complete response to neoadjuvant chemoradiotherapy (nCRT) in esophageal squamous cell carcinoma (ESCC).

Materials and methods: Patients were retrospectively enrolled between April 2007 and December 2018 from two institutions. We extracted deep learning features of six pre-trained convolutional neural networks, respectively, from pretreatment CT images in the training cohort (n = 161). Support vector machine was adopted as the classifier. Validation was performed in an external testing cohort (n = 70). We assessed the performance using the area under the receiver operating characteristics curve (AUC) and selected an optimal model, which was compared with a radiomics model developed from the training cohort. A clinical model consisting of clinical factors only was also built for baseline comparison. We further conducted a radiogenomics analysis using gene expression profiles to reveal underlying biology associated with radiological prediction.

Results: The optimal model with features extracted from ResNet50 achieved an AUC and accuracy of 0.805 (95% CI, 0.696–0.913) and 77.1% (65.6%-86.3%) in the testing cohort, compared with 0.725 (0.605–0.846)) and 67.1% (54.9%-77.9%) for the radiomics model. All the radiological models showed better predictive performance than the clinical model. Radiogenomics analysis suggested a potential association mainly with WNT signaling pathway and tumor microenvironment.

Conclusions: The novel and noninvasive deep learning approach could provide efficient and accurate prediction of treatment response to nCRT in ESCC, and benefit clinical decision making of therapeutic strategy.

Neoadjuvant chemoradiotherapy (nCRT) plus surgery has been shown to improve long-term outcomes in locally advanced esophageal squamous cell carcinoma (ESCC), especially those who achieve pathologic complete response (pCR) [1,2]. However, only 33–49% obtain pCR due to tumor heterogeneity [1–3]. Therefore, the pretreatment evaluation of response significantly affects the implementation of nCRT.

The development of machine learning and artificial intelligence have provided a new scope of computed tomography (CT) imaging analysis. Convolutional neural networks (CNNs) have been shown to improve diagnostic accuracy of medical imaging [4]. Because of the inherent limitation of sample size, training a CNN model from scratch for one specific clinical question often does not yield satisfactory results. An effective method is to employ transfer learning using pre-trained CNNs, which is frequently used to overcome the limitation of small data sets [5]. Part of traditional imaging descriptors developed for natural object detection have been

* Corresponding authors at: Department of Thoracic Surgery, Sun Yat-sen University Cancer Center, 651 Dongfeng Road East, Guangzhou 510060, China (J. Fu). Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Room 406, Block K, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong, China (V. Vardhanabhuti).

E-mail addresses: varv@hku.hk (V. Vardhanabhuti), fu_jh@outlook.com (J. Fu).

[1] Both authors contributed equally to this work as co-first authors.

https://doi.org/10.1016/j.radonc.2020.09.014
0167-8140/

commonly implemented for lesion segmentation in medical imaging analyses [6]. An alternative is to use pre-trained CNNs as feature extractors and conventional machine learning methods as classifiers, which might also have satisfying performance in predictive accuracy and computational costs for some tasks [7]. Handcrafted radiomics has been studied widely for radiological diagnosis and prediction [8], while the application of deep learning to evaluate nCRT response in esophageal cancer has not been explored yet.

The integration of various omics data (such as genomics, proteomics, and radiomics) for knowledge discovery has drawn much attention. Many gene expression patterns could be reflected by imaging traits [9], which allows for the decoding of molecular characterization in a noninvasive manner.

In this context, we hypothesize that CT-based deep learning approaches can be exploited to evaluate response to nCRT in ESCC. To this end, our study aims to develop a noninvasive measurement based on deep learning methods to predict pCR and externally validate the predictive ability in an independent cohort, making comparison with the handcrafted radiomics approach. Additionally, in a subset of patients, we used exploratory radiogenomics analysis to discern the underlying biological mechanisms of the radiological prediction.

Materials and methods

Patients

We reviewed the records of ESCC patients receiving nCRT plus surgery from Sun Yat-sen University Cancer Center (cohort 1 for model training) and the University of Hong Kong (cohort 2 for validation) between April 2007 and December 2018. The selection criteria included: (a) patients aged 18–80 years; (b) had histologically confirmed ESCC; (c) had standardized baseline enhanced CT scans; and (d) received nCRT plus surgery. The exclusion criteria included: (a) patients who underwent anticancer treatments before the baseline CT scans; (b) with a history of other malignancies; and (c) with incomplete medical records. pCR was defined as no viable residual tumor cells within the operative specimens including the resected primary tumor and lymph nodes. The study was designed and conducted in accordance with the Declaration of Helsinki. The institutional review boards of both participating centers approved this study. Patient consent was waived due to the retrospective design. The regimen of nCRT plus surgery and CT data acquisition are detailed in Supplementary Methods.

Regions of interest (ROIs)

The ROIs of primary tumor were manually segmented in CT images by two experienced radiologists (V.V and H.L) using ITK-SNAP software. For radiomics feature extraction, in order to make a fair comparison with deep learning features, the contoured regions covered the whole tumor in three consecutive slices with the maximum cross-sectional area of the tumor lesion. The ROI covering the whole tumor volume were also contoured for comparison. We assessed the interobserver reproducibility for ROI-based radiomics features using the CT images of 30 patients who were randomly chosen from the training cohort in a blinded manner. Segmentation procedure was repeated after a week to assess the intraobserver reproducibility. Features with intraclass correlation coefficients > 0.8 were selected for further analysis. For deep learning feature extraction, the 3 axial slices containing the delineated tumor were resized to 224 × 224 mm (the size for the input layer of the pretrained CNN models) with the use of a bounding box covering the radiologist-contoured tumor area.

Deep learning features

We used Xception [10], VGG16 [11], VGG19 [11], ResNet50 [12], InceptionV3 [13], or InceptionResNetV2 [14] as the base models for the extraction of representational features. These six commonly used CNNs were pre-trained on the large-scale, well-annotated ImageNet database. We removed the last fully connected layer at the top of the network and used global max pooling to take the maximum values of each layer of the feature maps to transform feature maps to raw values. These extracted features were then input into the steps of machine-learning model construction. The underlying mechanism for the prediction value of deep learning is not clear because of the complexity in the model structure. Additional details are provided in Supplementary Methods. Guided Gradient-weighted Class Activation Mapping (Guided Grad-CAM) [15] visualizes the CNN output in the last convolutional layer. We used this tool to investigate which locations were important in the derived deep learning features.

Handcrafted radiomics features

Handcrafted radiomics features were computed from the radiologist-drawn ROIs using the PyRadiomics package (version 2.1.2) in an automated manner [16]. Defined radiomics features with or without wavelet filtration were extracted. The study design was based on the image biomarker standardization initiative (IBSI) reporting guidelines [17]. Biomarker sets can be divided into 3 groups: (I) first-order statistics; (II) shape features; and (III) second-order features: gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), neighborhood gray tone difference matrix (NGTDM). Most features mentioned above were in accordance with feature definitions as described by the IBSI [18,19] with additional details specified in Supplementary Methods.

Harmonization

Radiological features are easily affected by CT acquisition and reconstruction parameters. The standardization of platforms and parameters from different institutions in advance is impractical in clinical settings. Among the methods developed to address the batch effect, ComBat [20] harmonization has been widely implemented in genomic studies and has been recently shown to correct the difference in radiological feature values resulting from different image acquisition protocols [21]. We harmonized deep learning and radiomics features after extraction.

Feature selection and model construction

We selected deep learning or radiomics features based on the following steps in the training cohort. The top 20% best features predictive of a pCR calculated by univariate analysis were selected. Then, we used a wrapper feature selection method based on the recursive feature addition algorithm to select the most predictive features that were evaluated by the area under the receiver operating characteristic curve (AUC). Classification was performed by the support vector machine (SVM) using a radial basis kernel function [22]. Clinical model classification probability was regarded as the radiological score. The code for model construction is available on Github (https://github.com/chenyixie123/ESCC_ML).

Radiogenomics association analysis and pathway enrichment analysis

The genomic profiles of pretreatment tissue specimens from 28 primary tumors of ESCC patients (11 pCR and 17 non-pCR) in

cohort 1 were measured using a GeneChip® Human Genome U133 Plus 2.0 Array [23]. The accession number of the gene expression data is GSE45670. We conducted spearman rank correlation tests to explore the relationship between gene expression and prediction probability. Correlated genes were recognized if *P* values were <0.05 and ranked by *P* values. This ranked gene list was used for pathway enrichment analysis using g:Profiler [24]. Biological process of the Gene Ontology (GO) was tested. The GO term size was set between 15 and 500. The threshold of the false discovery rate (*Q* value) was 0.05. An enrichment network was generated by a Cytoscape application, EnrichmentMap [25,26]. The visualization and clustering procedures followed a previously published protocol [27]. Finally, we used the AutoAnnotate application [28] to add summarized annotations to every cluster.

Statistical analysis

P values for differences in the clinical characteristics between cohorts were calculated by Fisher's exact test or Chi-square test for categorical data and Kruskal-Wallis test for numeric data. The AUC of the receiver operating characteristic (ROC) curve was adopted to determine the prediction performance. The calibration performance was measured quantitatively by the Hosmer-Lemeshow test and graphically by calibration plots [29]. A two-tailed *P* value <0.05 was defined as statistical significance. We performed statistical analysis and graphic production using Python v3.7 and R v3.3.1. The packages used in this study are shown in Supplementary Methods.

Results

Fig. 1 depicts the workflow processes. Of the 231 patients (mean age: 60 years; male: 83.1%) with locally advanced ESCC eligible for this study, 161 patients from cohort 1 were assigned to the training cohort, and 70 patients from cohort 2 were assigned to the external testing cohort. Clinical characteristics are shown in Table 1. No significant difference in pCR rate was identified (46.0% vs. 44.3%, *P* = 0.93) between the two cohorts. No clinical characteristics were significantly predictive for pCR (Table S1). BMI, smoking status, overall staging, sex, tumor location, histologic grade and family history of tumor were selected for the clinical model predicting pCR, yielding an AUC of 0.780 and 0.508 for the training and testing cohorts, respectively (Table 2, S2).

For handcrafted radiomics model construction, 851 features were extracted, where 107 and 744 were from original and wavelet filtered images, respectively. The imaging feature distributions differed between cohorts but overlapped better after ComBat harmonization (Fig. S1). Seven features were selected, including one from original and six from wavelet filtered images (Table S3). The radiomics model achieved an AUC of 0.822 in the training cohort, and an AUC of 0.725, C-index of 0.725, accuracy of 67.1%, sensitivity of 80.6%, specificity of 56.4%, positive predictive value (PPV) of 59.5%, and negative predictive value (NPV) of 78.6% in the testing cohort (Table 2, S2). We also developed a radiomics model using the same procedures based on features extracted from the whole contoured volume, and found that the performance remained similar with the one using three consecutive slices with maximum lesion for extraction (Table 2, S2 and S3).

As for predicting pCR using deep learning features, we compared six models adopting different CNNs as feature extractors to optimize prediction performance. The AUC ranged from 0.807 to 0.901 for the training cohort, and 0.635 to 0.805 for the testing cohort (Table 2, S2). The model using ResNet50 (RN-SVM model) contained 14 features, and achieved the best classification performance among the six and was superior to the radiomics model, yielding an AUC of 0.805, C-index of 0.805, accuracy of 77.1%, sensitivity of 83.9%, specificity of 71.8%, PPV of 70.3%, and NPV of 84.8% in the testing cohort (Fig. 2a, b). The model also showed good calibration and favorable clinical benefit (Fig. 2c, d). The number of features used in the models for other CNNs are summarized in Table S4. Feature maps generated from ResNet50 indicated locations that were important in generating the output (Fig. 3). Tumoral and peri-tumoral areas in the images were shown to be valuable for the feature pattern extraction. We also analyzed the performance generated by features extracted from earlier layers to see whether the last layer before the fully connected layer was the most suitable to extract features. As shown in Table S5, the current extraction strategy is the optimal one for ResNet50. Different feature selection methods and classifiers might greatly affect the predictive performance. We have compared cross combinations of multiple feature selection methods and classifiers for features extracted from different CNNs. The performance indeed varied among different combinations, and we found that the current method of combination of extraction and classifier we used was the best for our dataset (Fig. S2, Table S6), but generalizability to other datasets will need to be tested in the future.

We further integrated the deep learning and radiomics features to explore whether the predictive capability could be improved. The combination of deep learning and radiomics features failed to show a better performance, with a comparable AUC of 0.799 in the testing cohort (Fig. S3). We also evaluated the addition of clinical factors to radiological features for potential improvement of performance. Deep and/or radiomics features with clinical factors were incorporated into the machine learning model construction workflow, and clinical features were not selected for the model construction, indicating that the combination of clinical factors into radiological models could not increase the prediction performance.

The RN-SVM model was adopted to generate radiological scores for further radiogenomics analysis. A ranked gene list containing 726 genes with expression significantly correlated with radiological scores was used to perform pathway enrichment analysis. Among 385 enriched gene sets, biological processes involving the extracellular matrix (ECM) and WNT signaling pathway were predominant among the top ten candidates according to *Q* values (Table S7). After clustering highly interrelated gene sets, we focused on the top 20 clusters with larger sizes and more significant median *Q* values (Fig. 4, Fig. S4). Gene sets relating to WNT pathways made up the largest cluster in the network. Other signal transduction pathways such as the transforming growth factor β (TGF-β) pathway and peptide hormone signaling, were also among the most selected. Many microenvironmental components were involved, including proteoglycan, ECM, immune cells and hypoxia. Notably, biological procedures directly associated with radiotherapy and antimitotic chemotherapy, such as response to radiation and mitotic nuclear division, were also significantly enriched.

Discussion

The study developed prediction models for nCRT response in ESCC based on the deep learning or the handcrafted radiomics methods respectively, and performed validation in an independent cohort. We adopted transfer learning technique and extracted deep learning features from various pre-trained CNNs. RN-SVM model was the optimal one and had a better performance than the handcrafted radiomics model.

As an emerging image quantification method, radiomics has been investigated for its potential to predict nCRT response in esophageal cancer. Previous studies mainly focused on the application of PET-based radiomics [30–32], while only Yang's study [33]

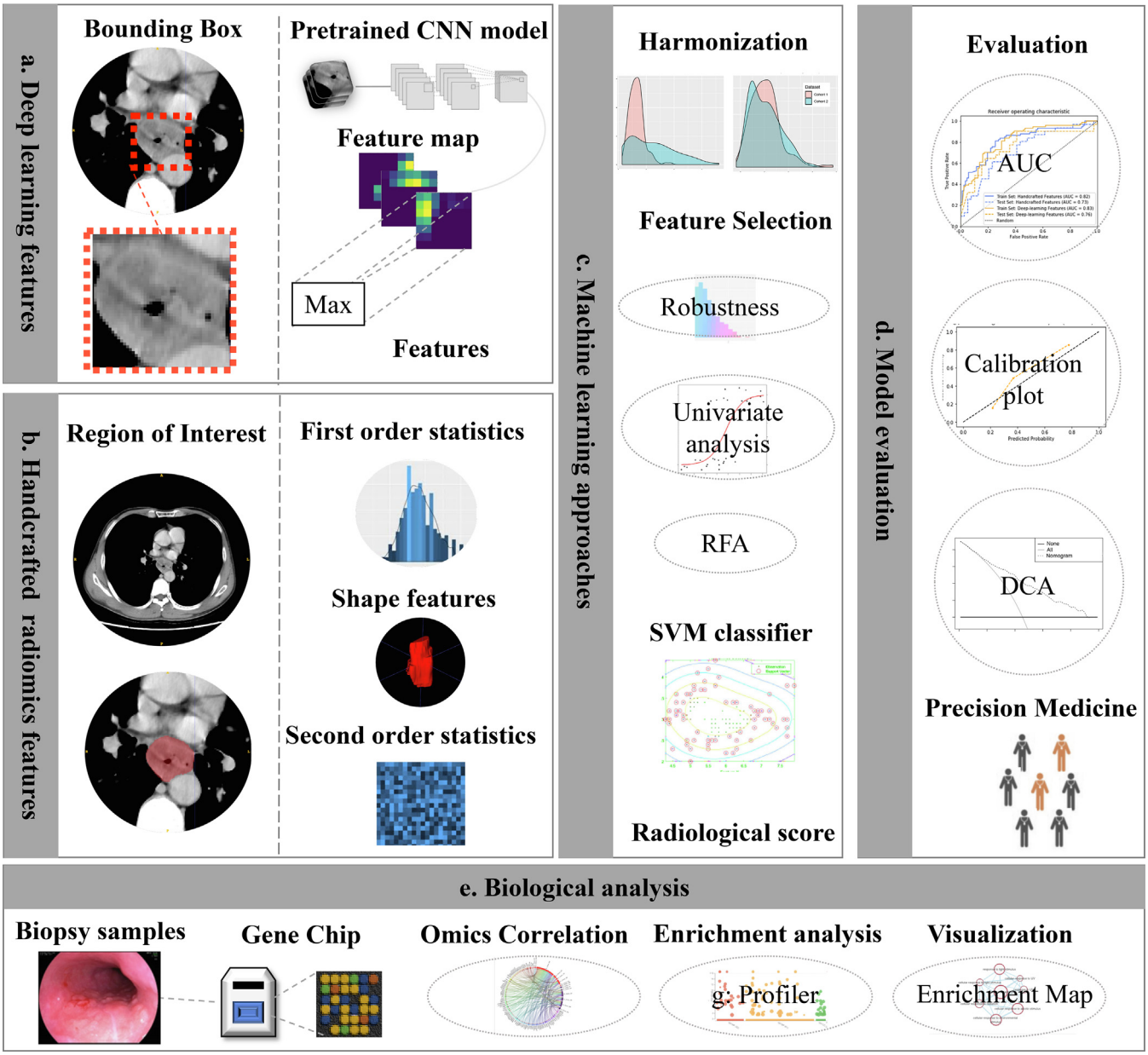

Fig. 1. Analysis flowchart. (a, b) Radiological models were constructed with the selected features extracted from the deep learning method and handcrafted radiomics method. (c) Machine learning methods were employed in model construction. (d) Model evaluation. (e) Radiogenomics analysis. CNN, convolutional neural network; RFA, recursive feature addition; SVM, support vector machine; AUC, area under the receiver operating characteristic curve; DCA, decision curve analysis.

employed CT-based features. The largest cohort containing 73 participants was used in Beukinga's study [31] to develop radiomics model, but adenocarcinoma accounted for the majority (89.0%) in this study. These researches characterized relatively small sample sizes and did not explore the extrapolation of developed models by external validation. Moreover, deep learning has not been used in the field of radiological prediction of nCRT response in esophageal cancer. The current study focused on squamous cell carcinoma, the histopathological type accounting for most of the cases worldwide. We established a CT-based model using the novel deep learning technique with a relatively larger sample size on a multicenter scale, and validated the model's generalizability. Moreover, deep learning feature extraction only needs the procedure of setting a bounding box of fixed size to the tumor region, which not only enhances efficiency but also reduces the variation of manual segmentation in the radiomics procedure, facilitating its generalizability.

Recent advances in deep learning as applied to medical imaging are expected to have significant contributions. However, training a deep learning neural network from scratch is prone to overfitting because the number of labeled medical images for one specific clinical question is typically limited. The use of a pre-trained CNN as a feature extractor in the medical domain has been proposed as one effective way to solve these difficulties [5,34–36]. Transfer learning can transfer prior knowledge of image characteristics and apply it to medical imaging, which has advantages in terms of better generalizability and ease of replication and testing. Our study demonstrated deep learning features extracted by the transfer learning method generalized well in medical tasks and achieved reasonably good results.

Table 1
Patient characteristics in the training and testing cohorts.

Characteristic	Training cohort (*N* = 161)	Testing cohort (*N* = 70)	*P* value
pCR			0.93
No	87 (54.0%)	39 (55.7%)	
Yes	74 (46.0%)	31 (44.3%)	
Sex			1.00
Male	134 (83.2%)	58 (82.9%)	
Female	27 (16.8%)	12 (17.1%)	
Age, years (mean [SD])	57.95 (6.86)	64.14 (10.65)	<0.01
cT stage			<0.01
1b	1 (0.6%)	0 (0)	
2	42 (26.1%)	3 (4.3%)	
3	114 (70.8%)	66 (94.3%)	
4a	4 (2.5%)	1 (1.4%)	
cN stage			0.48
0	8 (5.0%)	5 (7.1%)	
1	79 (49.1%)	27 (38.6%)	
2	60 (37.3%)	32 (45.7%)	
3	14 (8.7%)	6 (8.6%)	
Clinical stage			0.22
I	1 (0.6%)	0 (0)	
II	26 (16.1%)	5 (7.1%)	
III	115 (71.4%)	58 (82.9%)	
IV A	19 (11.8%)	7 (10.0%)	
Tumor location			<0.01
Upper	17 (10.6%)	2 (2.9%)	
Middle	95 (59.0%)	31 (44.3%)	
Lower	49 (30.4%)	37 (52.9%)	
Tumor length, cm (mean [SD])	5.50 (2.05)	5.76 (2.06)	0.39
Histologic grade			0.73
1	8 (5.0%)	3 (4.3%)	
2	104 (64.6%)	49 (70.0%)	
3	49 (30.4%)	18 (25.7%)	

Abbreviations: pCR, pathologic complete response.

The feature maps generated from CNN provided additional information on the important areas in the images for feature generation, which indicated the ability of deep learning to discover spatial heterogeneity of a tumor. The marginal area corresponding to the tumoral microenvironment and attached tissues might be useful for image pattern identification (Fig. 3).

The combination of handcrafted radiomics and deep learning features did not enhance the prediction performance in our study (Fig. S3), which was similar to the results published by Yun [35]. Hosny *et al.* [37] commented that the combination of radiomics and deep learning might incorporate human biases. We prefer the view that computationally derived imaging features from different frameworks could have different advanced-dimensional characteristics, which are undiscernible on the macroscopic scale. Human-defined radiomics might not fully encapsulate the differences between types of tissues.

The radiogenomics analysis highlighted the potential important roles of several signaling pathways and tumor microenvironment in the response assessment. Previous studies exploring mechanisms of conventional treatment resistance have revealed involvement of the WNT and TGF-β pathways [38,39], and microenvironmental components including ECM, proteoglycan, immune cells and hypoxia [40–42].

Our study has some limitations that are worth noting. First, although our results showed good prediction performance, indicating that the domain differences could be mitigated by transfer learning, heterogeneity existed between the source and target databases. One major obstacle in this research area is the development of a large public database with sufficient amounts of annotated medical imaging data to train plenty parameters in the neural network. Such database will dramatically help to provide more clinically relevant features to train model with better performance. Second, for the feature repeatability, due to the retrospective nature of this study, we adopted the ROI contour-based method but not test–retest imaging [17,19,43] for the evaluation of feature robustness. Test-retest study needs to be performed in a prospective manner in the future. Perturbation methods could be an alternative option [18] but appropriate method specific for ESCC has not yet been investigated. Third, we used Combat harmonization to minimize scanner-to scanner variability in order to address the batch effect. However, application of the trained model to new data of a patient sample potentially from different center needs to incorporate new data for training prior to Combat transformation. Further studies with standardization across a range of scanner parameters could help to harmonize the image features beforehand. Last, some enriched clusters in the radiogenomics analysis failed to show links to treatment response based on previous knowledge. Further studies should be conducted to validate the associations and confirm whether novel mechanisms exist.

In conclusion, our study developed and validated a model using transfer learning technique to perform pre-therapeutic assessment of nCRT response in ESCC. The RN-SVM model offers the advantages of better performance and a more efficient procedure

Table 2
Predictive performance of radiological or clinical models in the testing cohort.

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Xception-SVM	0.763 (0.643, 0.884)	71.4 (59.4, 81.6)	71.0 (55.0, 86.9)	71.8 (57.7, 85.9)	66.7 (50.6, 82.8)	75.7 (61.9, 89.5)
VGG16-SVM	0.648 (0.514, 0.781)	61.4 (49.0, 72.8)	67.7 (51.3, 84.2)	56.4 (40.8, 72.0)	55.3 (39.5, 71.1)	68.8 (52.7, 84.8)
VGG19-SVM	0.635 (0.499, 0.771)	61.4 (49.0, 72.8)	71.0 (55.0, 86.9)	53.8 (38.2, 69.5)	55.0 (39.6, 90.4)	70.0 (53.6, 86.4)
ResNet50-SVM	0.805 (0.696, 0.913)	77.1 (65.6, 86.3)	83.9 (70.9, 96.8)	71.8 (57.7, 85.9)	70.3 (55.5, 85.0)	84.8 (72.6, 97.1)
InceptionV3-SVM	0.753 (0.638, 0.867)	68.6 (56.4, 79.2)	54.8 (37.3, 72.4)	79.5 (66.8, 92.2)	68.0 (49.7, 86.3)	68.9 (55.4, 82.4)
InceptionResNetV2-SVM	0.653 (0.522, 0.783)	65.7 (53.4, 76.7)	74.2 (58.8, 89.6)	59.0 (43.5, 74.4)	59.0 (43.5, 74.4)	74.2 (58.8, 89.6)
Radiomics (3-slice)	0.725 (0.605, 0.846)	67.1 (54.9, 77.9)	80.6 (66.7, 94.6)	56.4 (40.8, 72.0)	59.5 (44.7, 74.4)	78.6 (63.4, 93.8)
Radiomics (whole volume)	0.712 (0.589, 0.836)	64.3 (51.9 75.4)	71.0 (55.0, 86.9)	59.0 (43.5, 74.4)	57.9 (42.2, 73.6)	71.9 (56.3, 87.5)
Clinical signature	0.508 (0.368, 0.649)	47.1 (35.1, 59.5)	54.8 (37.3, 72.4)	41.0 (25.6, 56.4)	42.5 (27.2, 57.8)	53.3 (35.5, 71.2)

Data are presented as percentages except AUC; 95% confidence intervals are included in parentheses.
Abbreviations: AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.

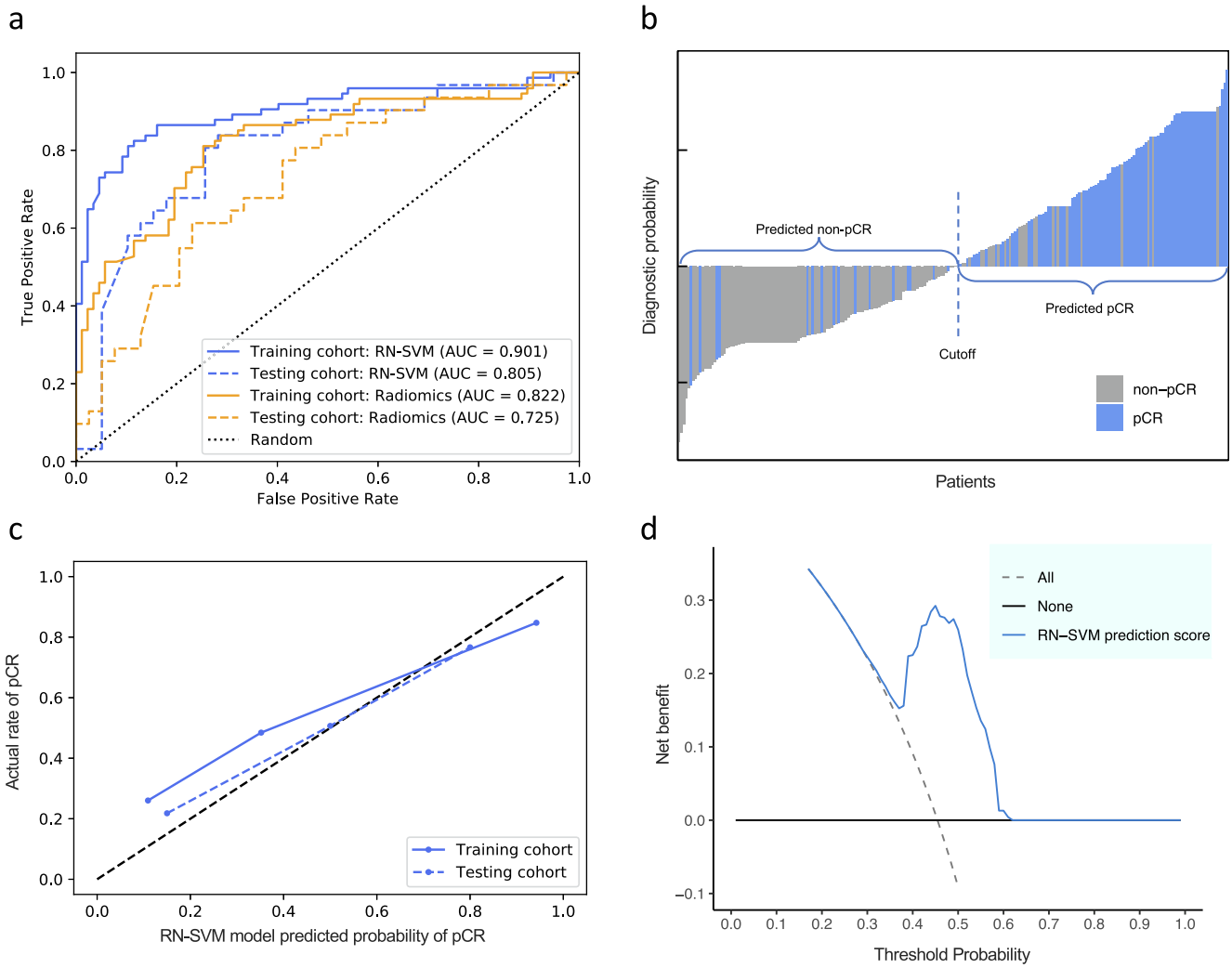

Fig. 2. Evaluation of predictive performances for the RN-SVM model and the radiomics model. (a) The receiver operating characteristic curves showing the predictive performances of the RN-SVM model and the handcrafted radiomics model in the training and testing cohorts, respectively. (b) Pathologic response status of patients in the pCR group and non-pCR group predicted by the RN-SVM model. The threshold was 0.448 determined by Youden's Index. (c, d) Curves of the calibration analysis and the decision curve analysis for the RN-SVM model. DL, deep learning; AUC, area under the receiver operating characteristic curve; pCR, pathologic complete response.

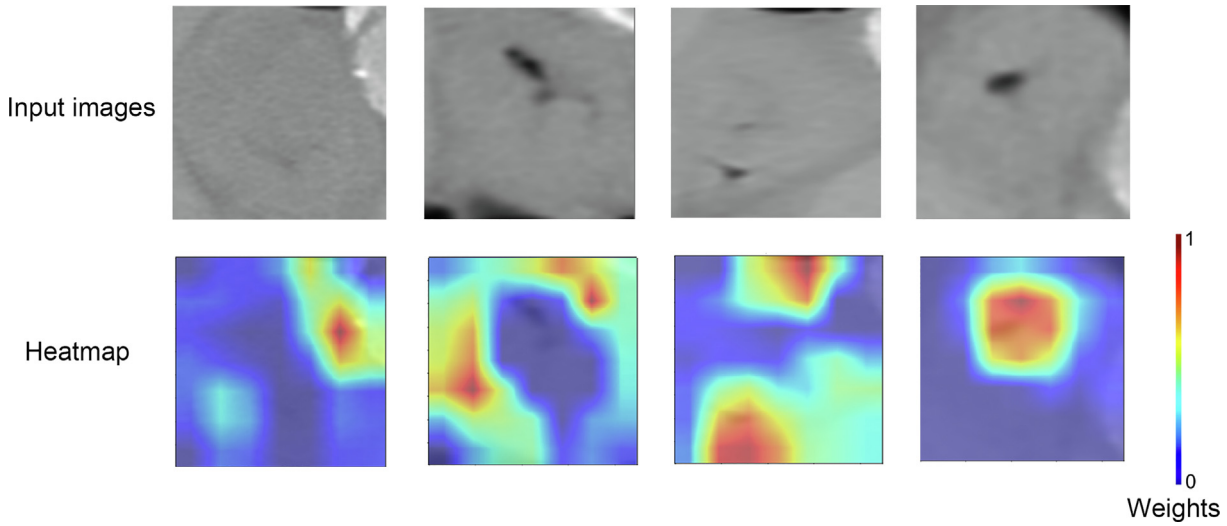

Fig. 3. Feature heatmaps of representative patients generated from the ResNet50 based on the Guided Grad-CAM. Gray-scale CT images and the corresponding feature heatmaps. The scaled weights of deep learning features are represented by the color bar.

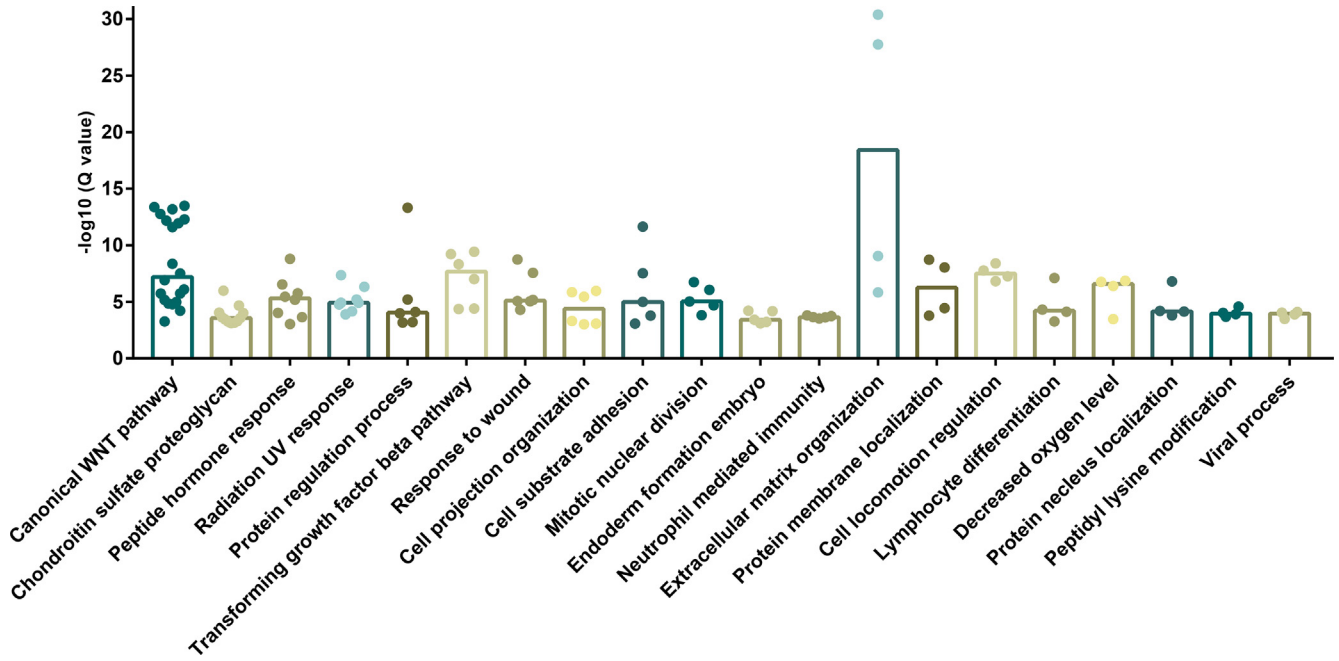

Fig. 4. Clusters of gene sets derived from pathway enrichment analysis. The scatter dot plot illustrated the top 20 clusters in the enrichment network, ranked by the cluster sizes and median *Q* values. Dots represent gene sets in a cluster. Bars represent medians of -log10 (*Q* values).

without manual tumor contouring compared with the handcrafted radiomics method. Transfer learning might be a good alternative for medical imaging tasks without sufficient data. Radiogenomics analysis provided useful insights into the biological mechanisms of conventional therapy resistance for esophageal cancer.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Health & Medical Collaborative Innovation Project of Guangzhou, China (grant number 201803040018); the National Natural Science Foundation of China, China (grant numbers 81972614 and 81871975); and the Fundamental Research Funds for the Central Universities, China (grant number 19ykyjs79). C.X. is supported by the Hui Pun Hing Memorial Postgraduate Fellowship from the University of Hong Kong, Hong Kong SAR, China. The funders had no role in study design, data collection and analysis, interpretation, or manuscript writing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2020.09.014.

References

[1] van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. N Engl J Med 2012;366:2074–84.
[2] Yang H, Liu H, Chen Y, Zhu C, Fang W, Yu Z, et al. Neoadjuvant chemoradiotherapy followed by surgery versus surgery alone for locally advanced squamous cell carcinoma of the esophagus (NEOCRTEC5010): A phase III multicenter, randomized, open-label clinical trial. J Clin Oncol 2018;36:2796–803.
[3] Mariette C, Dahan L, Mornex F, Maillard E, Thomas PA, Meunier B, et al. Surgery alone versus chemoradiotherapy followed by surgery for stage I and II esophageal cancer: final analysis of randomized controlled phase III trial FFCD 9901. J Clin Oncol 2014;32:2416–22.
[4] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018;172(1122–31):e9.
[5] Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35:1285–98.
[6] Lowe DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vision 2004;60:91–110.
[7] Raghu S, Sriraam N, Temel Y, Rao SV, Kubben PL. EEG based multi-class seizure type classification using convolutional neural network and transfer learning. Neural Netw 2020;124:202–12.
[8] van Rossum PSN, Xu C, Fried DV, Goense L, Court LE, Lin SH. The emerging field of radiomics in esophageal cancer: current evidence and future potential. Transl Cancer Res 2016;5:410–23.
[9] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[10] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:1800-7.
[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. CoRR 2014. abs/1409.1556.
[12] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:770-8.
[13] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015:2818-26.
[14] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI2016.
[15] Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vision 2016;128:336–59.
[16] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104–7.
[17] Balagurunathan Y, Kumar V, Gu Y, et al. Test-retest reproducibility analysis of lung CT image features. J Digit Imaging 2014;27(6):805–23.
[18] Zwanenburg A, Leger S, Agolli L, et al. Assessing robustness of radiomic features by image perturbation. Sci Rep 2019;9(1):614.
[19] Fiset S, Welch ML, Weiss J, et al. Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. Radiother Oncol 2019;135:107–14.
[20] Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, et al. Batch effect removal methods for microarray gene expression data integration: a survey. Brief Bioinform 2013;14:469–90.

[21] Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A method to compensate multicenter effects affecting CT radiomics. Radiology 2019;291:53–9.
[22] Amari S, Wu S. Improving support vector machine classifiers by modifying kernel functions. Neural Netw 1999;12:783–9.
[23] Wen J, Yang H, Liu MZ, Luo KJ, Liu H, Hu Y, et al. Gene expression analysis of pretreatment biopsies predicts the pathological response of esophageal squamous cell carcinomas to neo-chemoradiotherapy. Ann Oncol 2014;25:1769–74.
[24] Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 2019;47:W191–8.
[25] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[26] Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS ONE 2010;5:e13984.
[27] Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc 2019;14: 482–517.
[28] Kucera M, Isserlin R, Arkhangorodsky A, AutoAnnotate BGD. A cytoscape app for summarizing networks with semantic annotations. F1000Res 2016;5:1717.
[29] 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;8:53.
[30] Yuan H, Tong DK, Vardhanabhuti V, Law SY, Chiu KW, Khong PL. PET/CT in the evaluation of treatment response to neoadjuvant chemoradiotherapy and prognostication in patients with locally advanced esophageal squamous cell carcinoma. Nucl Med Commun 2016;37:947–55.
[31] Beukinga RJ, Hulshoff JB, Mul VEM, Noordzij W, Kats-Ugurlu G, Slart R, et al. Prediction of response to neoadjuvant chemotherapy and radiation therapy with baseline and restaging (18)F-FDG PET imaging biomarkers in patients with esophageal cancer. Radiology 2018;287:983–92.
[32] Chen YH, Lue KH, Chu SC, Chang BS, Wang LY, Liu DW, et al. Combining the radiomic features and traditional parameters of (18)F-FDG PET with clinical profiles to improve prognostic stratification in patients with esophageal squamous cell carcinoma treated with neoadjuvant chemoradiotherapy and surgery. Ann Nucl Med 2019;33:657–70.
[33] Yang Z, He B, Zhuang X, Gao X, Wang D, Li M, et al. CT-based radiomic signatures for prediction of pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant chemoradiotherapy. J Radiat Res 2019;60:538–45.
[34] Lopes UK, Valiati JF. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput Biol Med 2017;89:135–43.
[35] Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. Sci Rep 2019;9:5746.
[36] Zhu Y, Man C, Gong L, Dong D, Yu X, Wang S, et al. A deep learning radiomics model for preoperative grading in meningioma. Eur J Radiol 2019;116:128–34.
[37] Hosny A, Aerts HJ, Mak RH. Handcrafted versus deep learning radiomics for prediction of cancer therapy response. Lancet Digital Health 2019;1:e106–7.
[38] Zhong Z, Virshup DM. Wnt signaling and drug resistance in cancer. Mol Pharmacol 2020;97:72–89.
[39] Elliott RL, Blobe GC. Role of transforming growth factor Beta in human cancer. J Clin Oncol 2005;23:2078–93.
[40] He F, Wang H, Li Y, Liu W, Gao X, Chen D, et al. SRPX2 knockdown inhibits cell proliferation and metastasis and promotes chemosensitivity in esophageal squamous cell carcinoma. Biomed Pharmacother 2019;109:671–8.
[41] Chargari C, Clemenson C, Martins I, Perfettini JL, Deutsch E. Understanding the functions of tumor stroma in resistance to ionizing radiation: emerging targets for pharmacological modulation. Drug Resist Updat 2013;16:10–21.
[42] Qu Y, Dou B, Tan H, Feng Y, Wang N, Wang D. Tumor microenvironment-driven non-cell-autonomous resistance to antineoplastic treatment. Mol Cancer 2019;18:69.
[43] Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J Nucl Med 2012;53(5):693–700.