89975_Auto_Edited.docx

WORD COUNT	5756	TIME SUBMITTED	25-JAN-2024 04:57PM
		PAPER ID	106417960

Manuscript NO: 89975

A nomogram based on multimodal magnetic resonance combined with B7-H3mRNA for preoperative lymph node prediction in esophagus cancer

Yan-Han Xu, Peng Lu, Ming-Cheng Gao, Rui Wang, Yang-Yang Li, Rong-Qi Guo, Wei-Song Zhang, Jian-Xiang Song

Abstract

Background: Accurate preoperative prediction of lymph node metastasis (LNM) in esophageal cancer (EC) patients is of crucial clinical significance for treatment planning and prognosis. Therefore, we aimed to develop a clinical radiomics nomogram that can predict the preoperative lymph node status in EC patients.

Methods: A total of 32 EC patients confirmed by clinical pathology (who underwent surgical treatment) were included. Real-time fluorescent quantitative reverse transcription-polymerase chain reaction (RT-PCR) was used to detect the expression of B7-H3 mRNA in esophageal cancer tissue obtained during preoperative gastroscopy, and its correlation with LNM was analyzed. Radiomics features were extracted from multi-modal magnetic resonance imaging (MRI) of EC using Pyradiomics in Python. Feature extraction, data dimensionality reduction and feature selection were performed using XGBoost model and leave-one-out cross-validation. Multivariable logistic regression analysis was used to establish the prediction model, which included radiomics features, LN status from CT reports, and B7-H3 mRNA expression, represented by a radiomics nomogram. Receiver operating characteristic area under the curve (AUC) and decision curve analysis (DCA) were used to evaluate the predictive performance and clinical application value of the model.

Results: The relative expression of B7-H3 mRNA in EC patients with lymph node metastasis was higher than in those without metastasis, and the difference was statistically significant (P<0.05). The AUC value in the ROC curve was 0.718 (95% CI 0.528-0.907), with a sensitivity of 0.733 and specificity of 0.706, indicating good diagnostic performance. The individualized clinical prediction nomogram included radiomics features, LN status from CT reports, and B7-H3 mRNA expression. The ROC curve demonstrated good diagnostic value, with an AUC value of 0.765 (95% CI 0.598-0.931), sensitivity of 0.800, and specificity of 0.706. Decision curve analysis indicated the practical value of the radiomics nomogram in clinical practice.

Conclusion: This study developed a radiomics nomogram that includes radiomics features, LN

status from CT reports, and B7-H3 mRNA expression, enabling convenient preoperative individualized prediction of lymph node metastasis in EC patients.

Keywords: Esophageal cancer; radiomics; B7-H3mRNA; multimodal magnetic resonance imaging; lymph node metastasis; nomogram

Core Tip: Accurate TNM staging plays a critical role in devising treatment strategies for esophageal cancer, particularly in assessing lymph node metastasis. Nevertheless, existing techniques for diagnosing lymph nodes in esophageal cancer are currently constrained by limited accuracy. In light of this, our study endeavors to construct a clinical column chart that can enhance the assessment of lymph node status, furnishing a valuable point of reference for the diagnosis and treatment of esophageal cancer.

1 Introduction

According to relevant research statistics, esophageal cancer is a common malignant tumor in the field of thoracic surgery, ranking seventh in terms of incidence and sixth in terms of mortality worldwide. In Asia, the main histological type is squamous cell carcinoma(1-3). Most esophageal cancer patients require comprehensive treatment. In the early stage, surgery or endoscopic resection is the primary approach, while concurrent chemoradiotherapy is preferred for patients in the middle and late stages. The specific treatment plan should be based on the accurate staging of esophageal cancer using the TNM classification system (4).

Simultaneously, in surgical treatment, due to the highly variable lymphatic spread of cancer, suspicious positive lymph nodes should be resected together with the tumor to improve patient survival. However, some studies suggest that expanding the range of lymph node dissection may increase postoperative complications and worsen prognosis for cancer patients(5, 6). Therefore, in the formulation of treatment strategies for esophageal cancer, accurate diagnosis of lymph node metastasis status is crucial(7, 8).

B7-H3, also known as CD276, is a member of the B7 ligand family and is an attractive target in antibody immunotherapy. It is overexpressed on many malignant cells and cancer stem cells but exhibits low-level expression in normal tissues(9). Relevant studies have shown that B7-H3 primarily promotes tumor development through immune mechanisms by inhibiting specific immune responses, leading to a pro-tumoral effect(9). Research has found associations between B7-H3 expression and clinical TNM progression and prognosis in diseases such as gastric cancer, pancreatic cancer, colorectal cancer, lung cancer, and acute myeloid leukemia(10-15). Additionally, Arigami et al. discovered a strong correlation between B7-H3 expression and sentinel lymph node metastasis and the number of lymph node metastases in their study of breast cancer. In their

multivariate analysis, the mRNA expression of B7-H3 in the primary tumor significantly predicted regional lymph node metastasis(16). Furthermore, Chen et al. found a close association between B7-H3 expression in esophageal cancer and aggressive biology, low tumor-infiltrating T lymphocyte density, and poor prognosis(17). However, there is currently no relevant research proving an association between B7-H3 expression and lymph node metastasis in esophageal cancer.

Currently, computed tomography (CT) is commonly used to determine the preoperative lymph node status in EC patients, primarily relying on size-based measurements (e.g., a 10 mm short-axis diameter on CT as the cutoff value for diagnosing lymph node metastasis)(18). However, relevant studies have shown that this method has an accuracy rate of less than 70% in determining lymph node metastasis(19). Positron emission tomography/CT (PET/CT) is a rapidly developing imaging modality that combines positron emission tomography with X-ray computed tomography. However, its application in lymph node metastasis diagnosis is limited due to its high cost, low sensitivity, and high false-positive rate(20). Meanwhile, research on positive lymph node detection in certain cancers suggests that magnetic resonance imaging (MRI) has higher accuracy(21, 22). However, conventional MRI of the chest is prone to motion artifacts due to respiratory motion, which can affect image quality. With the emergence of multi-sequence MRI techniques such as StarVIBE and T2TSE-BLADE, respiratory motion artifacts in non-breath-hold patients and uncooperative patients have been significantly reduced, resulting in clearer visualization of tumors and the surrounding soft tissue boundaries and improved image quality(23, 24). However, existing imaging modalities primarily rely on lymph node anatomy, and their assessment of lymph node metastasis is based on size measurements, which are insufficient to reveal the internal structural characteristics of lymph nodes and obtain valuable tumor-related information(25).

Radiomics research involves applying computer mathematical tools to image processing, extracting radiological features such as shape, texture, or waveform, which can provide information about cancer phenotypes and the tumor microenvironment. This concept was introduced by Lambin et al. in 2012(26). By combining radiomics-derived data with other relevant data, accurate and reliable Clinical Decision Support Systems (CDSS) can be generated(27). Currently, radiomics has made significant progress in the qualitative assessment of tumors, diagnosis of lymph node metastasis, and prognosis prediction(28-30).However, there is currently no research that combines radiomics with expression factors in primary tumors to elucidate their diagnostic and predictive value in cancer. This integration could contribute to a more comprehensive understanding of the biological characteristics and behavior of tumors, providing more accurate predictions for individualized treatment.

Therefore, the objective of this study is to develop a radiomics nomogram that combines radiomic features, B7-H3 mRNA expression levels, and clinical risk factors for individualized prediction of preoperative lymph node metastasis in EC patients.

2 Materials and Methods

2.1 Patients

Our research institution's ethics review committee (the Medical Ethics Committee of the Sixth Affiliated Hospital of Nantong University, Yancheng Third People's Hospital) has approved this research project. Considering that the relevant examinations in this study do not pose significant physical or harm to the patients' interests, the requirement for obtaining informed consent from the patients has been waived by the committee. Our study included a total of 32 esophageal cancer patients (9 females and 23 males) who received treatment at our hospital from March 2022 to July 2023 and met the inclusion criteria of this study. The patients had an average age of 70.53±6.41 years, with an age range of 52-84 years. The inclusion criteria were as follows: (1) all patients were over 18 years of age; (2) standard contrast-enhanced CT and MRI examinations were performed within 10 days before treatment; (3) patients underwent gastroscopy and pathological biopsy at our hospital, and the pathological diagnosis was confirmed as esophageal squamous cell carcinoma; (4) the surgical approach was consistent for all patients, with three-field lymph node dissection for esophageal cancer; (5) MRI images had sufficient clarity to support radiomic feature extraction; (6) availability of clinical and pathological information; (7) B7-H3 mRNA expression was determined by RT-PCR using cancer tissue samples obtained from preoperative esophagoscopic biopsies. The exclusion criteria were as follows: (1) patients with significant surgical contraindications; (2) patients with concurrent other tumor diseases; (3) inability to undergo MRI examination or presence of contraindications for MRI examination.

Baseline clinical information and pathological data of the patients in the study, including age, gender, tumor location, tumor size, and lymph node status (based on pathological results), were obtained from medical records. In addition, the enhanced CT reports of the study patients were collected, and lymph nodes with a size of ≥10.0mm in the CT reports were considered as positive for lymph node involvement.

2.2 Image acquisition and segmentation

The patients were scanned using a 3.0T magnetic resonance imaging (MRI) scanner (MAGNETOM Skyra 3.0T, Siemens Healthcare, Germany) and an 18-channel surface phased-array coil. Prior to the examination, patients were instructed to remove any metallic objects and undergo respiratory training. The patients were positioned in a supine position with the head first, and the scanning range extended from the bilateral lung apices to 1 cm below the diaphragm. The MRI scanning sequences included T1-Star-VIBE and T2-TSE-BLADE sequences. The parameters for the T1-Star-VIBE sequence were as follows: TR/TE = 3.98ms/1.91ms; voxel size = 1.0 mm × 1.0 mm × 1.0 mm; FOV = 300mm × 300mm; flip angle = 12°; scanning time = 309 seconds. The parameters for the T2-TSE-BLADE sequence were as follows: TR/TE = 5000ms/97ms; voxel size = 0.9 mm × 0.9 mm × 3.0 mm; FOV = 260 mm × 260 mm; flip angle = 180°; scanning time = 360-600 seconds.

We retrieved the MRI images of all patients from the hospital's Picture Archiving and Communication System (PACS). Preprocessing of the acquired images was performed using Python, which included bias field correction utilizing the N4 correction algorithm and registration alignment. Subsequently, image feature segmentation and analysis were conducted. These processes aimed to extract meaningful features from the images, facilitating further analysis and

the development of our study's models.

The three-dimensional (3D) semi-automatic segmentation was performed by a single operator, who was a thoracic surgery graduate student, using the 3D Slicer software. The segmentation process involved the extraction of valuable regions of interest (ROIs) based on the radiologist's interpretation. The radiologist, with 10 years of experience in the field, provided the expert assessment of the images, and the operator utilized this information to guide the segmentation and extraction of the ROIs. The operator carefully followed the radiologist's findings to ensure accurate and reliable segmentation of the desired regions.

In this study, ROIs referred to the primary lesions of esophageal cancer. These ROIs were manually delineated on each consecutive slice of the MRI images along the boundaries of the primary tumor lesions. The delineation process excluded adjacent air, blood vessels, fat, and normal tissues, focusing solely on the pathological features of the primary tumor. The delineation was performed by the operator using the 3D Slicer software, guided by the radiologist's interpretation and expertise.

2.3 Radiomics feature extraction and selection

We applied the PyRadiomics component in Python (https://pyradiomics.readthedocs.io/en/latest/) to extract features from the ROI for each case. Prior to ROI processing, we performed bias field correction (N4 correction) and registration alignment on every MRI image of all sequences to mitigate the impact of varying grayscale ranges in PyRadiomics. The extracted feature categories include first-order, shape, gray level co-occurrence matrix, gray level run length matrix, gray level size zone matrix, gray level dependence matrix, and neighboring gray tone difference matrix.

Due to the limited sample size in this study, we employed the XGBoost model and leave-one-out cross-validation method to construct the radiomics features. Firstly, a t-test was applied to select the top 30% features that are predictive of lymph node metastasis. Secondly, using the leave-one-out cross-validation method, internal cross-validation was performed to further retain features that improve diagnostic performance. Finally, the radiomics signature and its corresponding weight values were computed to obtain the consistency features in the model. The specific formula for the established and extracted radiomics features is as follows:

Radiomics signature = Intercept + coef1 × feature1 + coef2 × feature2 + coef3 × feature3 + coef4 × feature4 + coef5 × feature5 + ... + coefn × featuren

2.4 Real time quantitative reverse transcription polymerase chain reaction (RT-PCR)

2.4.1 Acquisition of Esophageal Pathological Tissue under Esophagoscopy

Patients diagnosed with esophageal cancer were carefully screened for inclusion in the study. Under gastroscopy, specialized endoscopic forceps were used to obtain biopsy specimens from suspected cancerous lesions. Additionally, samples of normal tissue were collected from a location at least 5 cm away from the suspected cancerous area to serve as adjacent normal tissue. Following the collection of cancerous and adjacent normal tissue, the specimens were immediately immersed and washed in physiological saline solution. Subsequently, they were stored at a temperature of -80°C in a freezer for preservation. To ensure the accuracy and reliability of the study, specimens that did not meet the predefined inclusion criteria were carefully

excluded. This exclusion process was based on a combination of subsequent patient treatment and pathological diagnosis, which served as the gold standard.

2.4.2 RNA Extraction from Specimen Tissue

2.4.2.1 Tissue Lysis

The collected specimens of esophageal cancer and adjacent normal tissue were removed from the -80°C freezer and thawed. On a sterile bench, the tissue was finely minced using sterile tissue scissors. Approximately 0.2g of the tissue was weighed on an electronic precision balance and transferred into a grinding tube. The tissue was then washed with PBS buffer (dissolving protective reagent) for 5 minutes. The grinding tube containing the tissue was placed on ice.

Next, 500μL of RNA extraction solution (e.g., RNA Extrizol) was added to the grinding tube using a pipette. A handheld tissue grinder was used, adjusting it to the maximum speed, to grind the tissue in a start-stop manner. Care was taken to avoid liquid splashing during the grinding process. Grinding was continued until the tissue was completely lysed and formed a homogenized liquid. Finally, 500μL of RNA extraction solution was added and mixed well with the lysate.

Subsequently, 200μL of chloroform solvent (trichloromethane) was added to the tube. The tube was then placed on a shaker and shaken for 15 seconds. After shaking, the tube was left to stand at room temperature for 3 minutes.

2.4.2.2 RNA Precipitation and Washing

The pre-cooled centrifuge was set to a temperature of 4°C. The centrifuge speed was adjusted to 13,000 rpm, and the centrifugation time was set to 15 minutes. The tubes containing the lysed tissue were placed in the centrifuge and balanced before initiating centrifugation. After 15 minutes of centrifugation, the top layer of liquid (approximately 400μL) from the tube was carefully extracted using a pipette and transferred to a new tube. It is important to only aspirate the top layer of liquid, avoiding any other layered liquids to prevent contamination of chromosomal DNA.

Using a pipette, 500μL of isopropanol was added to the tube containing the extracted liquid, and the mixture was thoroughly mixed. The tube was then left to stand at room temperature for 10 minutes to allow RNA precipitation.

The centrifuge process was repeated as described above (centrifuge temperature set at 4°C, centrifuge speed at 13,000 rpm, and centrifugation time of 15 minutes). The tube was placed in the centrifuge and balanced before initiating centrifugation. After 15 minutes of centrifugation, a white, flocculent precipitate could be observed at the bottom of the centrifuge tube, which represented the RNA after lysis completion. Using a pipette, the entire supernatant, excluding the flocculent precipitate, was aspirated and discarded, while retaining the RNA pellet.

To wash the RNA pellet, 1 mL of 75% ethanol solvent was added to the centrifuge tube and mixed thoroughly. The centrifuge parameters were set to a temperature of 4°C, a speed of 750 g, and a centrifugation time of 5 minutes. The tube was placed in the centrifuge, balanced, and then centrifuged.

2.4.2.3 RNA Resuspension and Concentration Measurement

Using a pipette, carefully aspirate the upper ethanol solution from the centrifuge tube, being cautious not to draw up the white precipitate at the bottom. After aspirating the ethanol solution, let the centrifuge tube air dry at room temperature. Then, add 20 μl of nuclease-free water to the dried centrifuge tube to fully dissolve the RNA.

Next, measure the concentration and purity of the RNA in the centrifuge tube using a NANODROP 2000 nucleic acid and protein analyzer. Ensure that the absorbance ratio at 260 nm and 280 nm falls within the range of 1.8-2.0 for the retained samples to ensure accuracy of the measurement results. If the ratio falls outside this range, discard the sample and repeat the experiment with a new tissue specimen. Label the centrifuge tube containing the remaining RNA and store it in a -80°C freezer for future experiments.

2.4.3 mRNA Real-time Quantitative PCR Experimental Steps

2.4.3.1 cDNA Synthesis by Reverse Transcription

RNA extraction solution was retrieved from a -80°C freezer and transferred to a new microcentrifuge tube. Using a pipette, 3μg of RNA was extracted as a template and mixed with 1μL of Oligo(dT) primer solution. Then, nuclease-free water was added to achieve a total volume of 12μL, and the mixture was gently mixed. The mixture was centrifuged for 10 seconds in a microcentrifuge.

Prior to the next step, a constant temperature incubator was pre-set at 70°C for 5 minutes. The microcentrifuge tube containing the mixture was heated in the incubator. After heating, the tube was immediately transferred to an icebox for cooling, followed by a brief centrifugation to collect the precipitate.

Next, the reaction mixture was prepared. Using a pipette, 4μL of 5× Reaction Buffer, 1μL of RNase Inhibitor, 2μL of 10mM dNTP Mix, and 1μL of RevertAid M-MuLV RT were drawn. Each microcentrifuge tube was then added with 8μL of the reaction mixture, gently mixed, and centrifuged for 5 seconds in a microcentrifuge.

Prior to performing the reverse transcription reaction, the PCR machine was pre-set to a temperature of 42°C, and a 6-minute wait was observed. Subsequently, the reaction was incubated at 72°C for 5 minutes to terminate the reverse transcription. The microcentrifuge tubes were placed in ice, and a 1:10 dilution was performed for storage at -20°C.

2.4.3.2 Real-time Quantitative PCR Amplification and Analysis

Retrieve an equal volume of cDNA template from the -20°C freezer and place it in an EP tube. Add the amplification primers to achieve a total volume of 20μL. Transfer the mixture to the PCR machine for further reaction.

Data analysis was performed using the 2-△CT relative quantification method, with β-actin as the reference gene for normalization. Each group included two technical replicates, and the procedure was as follows: Retrieve the standard EP tube and sequentially add 2μL of template, 1μL of upstream primer (concentration of 10μmol/L), 1μL of downstream primer (concentration of 10μmol/L), 12.5μL of SYBR Green Master Mix, and 8.5μL of double distilled water. Gently mix the solution and centrifuge for 10 seconds in a microcentrifuge.

Perform amplification using the PCR machine. Each group contains three samples, and the

program is set as follows: Pre-denaturation at 95°C for 10 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 40 seconds. Finally, terminate the reaction at 72°C for 10 minutes.

After the completion of the PCR amplification, data analysis was performed based on the amplification curve and melt curve. The 2-ΔΔCt relative quantification method was used for data analysis in each group, with β-actin as the internal reference gene for normalization. Each group included four samples, and three experimental replicates were performed. The final data analysis was conducted using GraphPad Prism 5.0 software.

2.5 Model construction

We utilized the R software to perform logistic regression analysis to identify independent predictive factors, including radiomic features, B7-H3 mRNA expression level, and LN status reported by CT, in our study. A model incorporating these independent predictors was developed, and its performance was evaluated using the area under the receiver operating characteristic curve (AUC).

2.6 Nomogram development and decision curve analysis (DCA)

Nomogram development and decision curve analysis (DCA) were employed for model visualization and clinical application. To assess the additional value of radiomic features, B7-H3 mRNA, and CT in individually predicting the preoperative lymph node (LN) status in esophageal cancer patients, four decision curves were developed based on CT reports, radiomic features of the primary lesion, B7-H3 mRNA, and a combined model (including CT reports, B7-H3 mRNA, and radiomic features). These decision curves were used to further determine the clinical utility of the plotted line chart.

3 Results

3.1 Clinical Characteristics

The study included a total of 32 esophageal cancer patients, as shown in Table 1. There were no significant differences ($P>0.05$) between the LN-positive and LN-negative groups in terms of patient age, gender, preoperative CEA levels, preoperative SCC levels, pathological grade, tumor location, and tumor size. However, a statistically significant difference was observed in T stage ($P<0.05$). In the study cohort, the lymph node metastasis rate was 46.88% (15/32) based on postoperative pathological diagnosis. Regarding the subjective enhanced CT reports of LN status, 6 patients were reported as LN-negative but confirmed to have LN metastasis, while 6 patients were reported as LN-positive but confirmed to be LN-negative. The sensitivity was 60.0%, and the specificity was 64.7%.

Table 1. Characteristics of Patients in the Cohort

Characteristic	LN Metastasis (+)	LN Metastasis (-)	P
Age, mean SD	70.18±6.75	71.92±7.49	0.39
Tumor size, mean SD	3.89±1.55	3.18±1.64	0.24
Sex(%)			0.54
Male	10 (66.67%)	13 (76.47%)	
Female	5 (33.33%)	4 (23.53%)	
CEA (%)			0.99
0–5 ng/mL	15 (99.99%)	16 (94.12%)	
> 5 ng/mL	0 (0.01%)	1 (5.88%)	
SCC (%)			0.61
0–2.7 ng/mL	12 (80.0%)	14 (82.35%)	
> 2.7 ng/mL	3 (20.00%)	3 (17.65%)	
Location (%)			0.99
Upper mediastinal	2 (13.33%)	1 (5.88%)	
Middle mediastinal	6 (40.00%)	8 (47.06%)	
Lower mediastinal	6 (40.00%)	8 (47.06%)	
Abdominal	1 (6.67%)	0 (0.01%)	
Histologic grade (%)			0.40
Well differentiated	0 (0.01%)	2 (11.77%)	
Moderately differentiated	11 (73.33%)	11 (64.71%)	
Poorly differentiated	4 (26.67%)	4 (23.53%)	
T stage(%)			0.01
1	3 (20.00%)	10 (58.82%)	
2	12 (80.00%)	5 (29.41%)	
3	0 (0.01%)	2 (11.77%)	
CT-reported LN status			0.16
0–10.0 mm	6 (40.00%)	11 (64.71%)	
> 10.0 mm	9 (60.00%)	6 (35.29%)	

3.2 The relationship between the expression level of B7-H3 mRNA and the clinical characteristics of esophageal cancer patients

The correlation between B7-H3 mRNA expression levels and age, gender, T stage, and N stage (lymph node metastasis) in patients with esophageal cancer was analyzed separately. The T stage and N stage of esophageal cancer were determined according to the International 8th edition TNM staging criteria for esophageal cancer.

Among the 32 patients with esophageal cancer, the average B7-H3 mRNA expression level in male patients was 2.47±0.51, while in female patients, it was 2.52±0.40. There was no significant statistical difference between the two groups ($P>0.05$). The results indicate that there is no significant correlation between B7-H3 mRNA expression levels and gender in patients with esophageal cancer ($P>0.05$).

Among the 32 patients with esophageal cancer, the average B7-H3 mRNA expression level in T1 stage patients was 2.23±0.32, in T2 stage patients it was 2.66±0.53, and in T3 stage patients it was 2.59±0.12. There was a significant correlation in the average B7-H3 mRNA expression levels among the different T stages ($P<0.05$).Furthermore, we found that the average B7-H3 mRNA expression level in the tissues of esophageal cancer patients with lymph node metastasis was 2.70±0.53, while in patients without lymph node metastasis, it was 2.29±0.35. The study revealed a significant increase in B7-H3 mRNA expression levels in the tissues of esophageal cancer patients with lymph node metastasis compared to those without lymph node metastasis ($P<0.05$) (Figure 2).

In conclusion, the expression level of B7-H3 mRNA in esophageal cancer is not significantly correlated with gender (P>0.05). However, it is closely associated with T stage and lymph node metastasis in esophageal cancer (P<0.05). The relative expression level of B7-H3 mRNA is significantly higher in lymph node-positive tissues of esophageal squamous cell carcinoma, indicating an upregulation of B7-H3 mRNA expression in lymph node-positive tissues of esophageal cancer.

Table 2. The relationship between B7-H3 expression and clinical characteristics of esophageal cancer patients

Clinical Characteristic	Number of sample	B7-H3mRNA	p
Sex			
male	23	2.47±0.51	0.80
female	9	2.52±0.40	
T			P
1	13	2.23±0.32	0.05
2	17	2.66±0.53	
3	2	2.59±0.12	
N			
0	17	2.29±0.35	0.02
1~3	15	2.70±0.53	

3.3 Diagnostic value of B7-H3 mRNA in detecting lymph node metastasis in esophageal cancer.

We performed quantitative analysis of B7-H3 mRNA expression in 32 esophageal cancer tissues and generated a receiver operating characteristic (ROC) curve (Figure 1) to evaluate the diagnostic value of B7-H3 in detecting lymph node metastasis in esophageal cancer using clinical pathological examination as the gold standard. The AUC for B7-H3 in detecting lymph node metastasis was 0.718, with a sensitivity of 73.3% and specificity of 70.6%. The optimal diagnostic threshold for B7-H3 in identifying lymph node metastasis in esophageal cancer was determined to be 2.56.

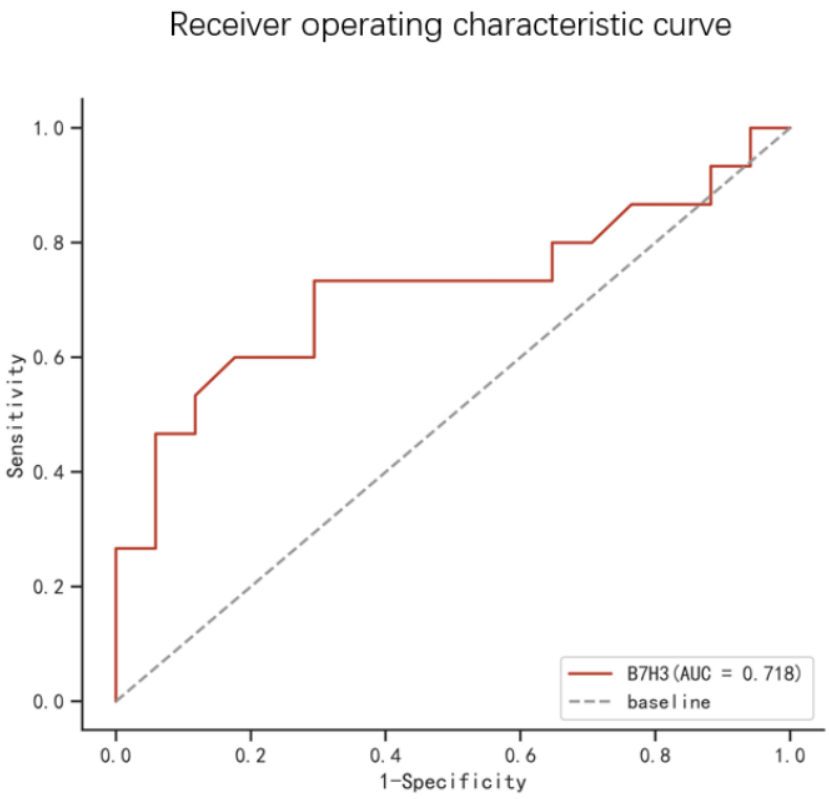


Figure 1 ROC curve of B7-H3mRNA in the study queue

3.4 Feature extraction and model construction

Using the Pyradiomics package in Python, a total of 1169 radiomic features were extracted from MRI images. The feature selection process was performed using the Leave-one-out method from the XGBoost model, resulting in the reduction of the feature set to 18 potential predictive factors (Figure 2A and 2B). In the LASSO regression model, the following features were identified as having non-zero coefficients:

T1-log-sigma-2-0-mm-3D_glcm_DifferenceAverage, T1-log-sigma-5-0-mm-3D_glszm_SizeZoneNonUniformityNormalized, T1-wavelet-LLH_gldm_LargeDependenceEmphasis, T1-wavelet-LHL_firstorder_Median, T1-wavelet-LHL_firstorder_RootMeanSquared, T1-wavelet-LHH_glcm_Contrast, T1-wavelet-LHH_glcm_JointEntropy, T1-wavelet-LHH_glszm_SmallAreaHighGrayLevelEmphasis, T1-wavelet-HLH_firstorder_MeanAbsoluteDeviation, T1-wavelet-HLH_glszm_SmallAreaEmphasis, T1-wavelet-HLH_glszm_SmallAreaHighGrayLevelEmphasis, T1-wavelet-HHL_glcm_ClusterTendency, T1-wavelet-HHL_glcm_Contrast, T1-wavelet-HHL_glszm_GrayLevelNonUniformity, T1-wavelet-HHH_firstorder_InterquartileRange, T2-original_gldm_DependenceEntropy, T2-log-sigma-4-0-mm-3D_glszm_GrayLevelNonUniformity, T2-wavelet-HLH_glcm_Idn. Additionally, the optimal weight value feature was determined to be T1-wavelet-HL

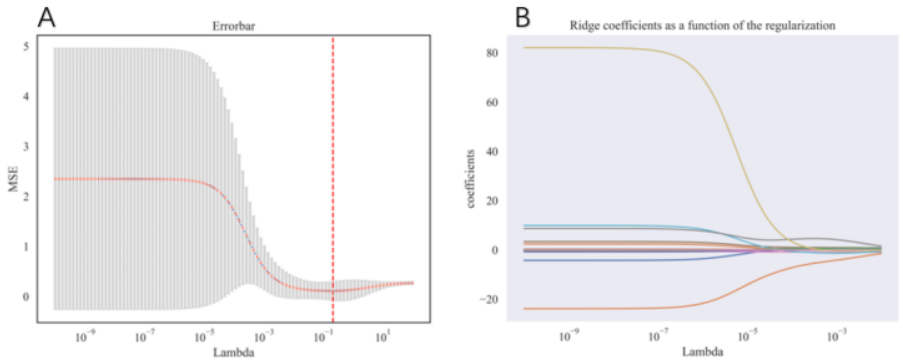
H_glcm_InverseVariance through further calculations and analysis.

Subsequently, logistic regression analysis was conducted in R language to further determine the independent predictive factors among the selected radiomic features, B7-H3 mRNA expression level, T stage, and LN status from CT reports. During the analysis, the model with the smallest Akaike information criterion (AIC) value was chosen.

The following radiomic features were further selected:

T1-wavelet-LHL_firstorder_Median, T1-wavelet-HLH_glszm_SmallAreaHighGrayLevelEmphasis, T1-wavelet-HHL_glcm_ClusterTendency, T1-wavelet-HHH_firstorder_InterquartileRange, T2-original_gldm_DependenceEntropy, T2-wavelet-HLH_glcm_Idn (the corresponding coefficient values for each independent predictive factor are detailed in Table 3).

A model incorporating these independent predictive factors was developed and presented in the form of a column chart (Figure 3).



7

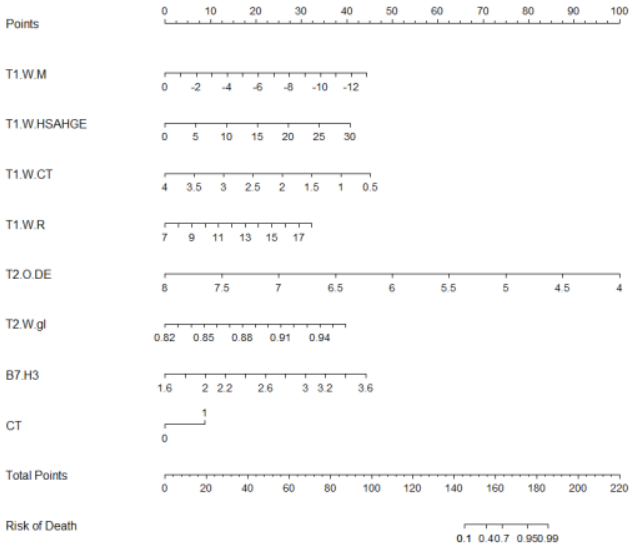
Figure 2: Texture feature selection using the Least Absolute Shrinkage and Selection Operator (LASSO) binary logistic regression model.

(A) The tuning parameter (Lambda) selection in the LASSO model was performed using 10-fold cross-validation with the minimum criterion. The relationship curve between the mean-square error (MSE) and Lambda is depicted, with a dashed line indicating the optimal value. The vertical lines represent the values selected through 10-fold cross-validation, including 18 optimized non-zero coefficients. (B) LASSO coefficient profiles of 1169 texture features. The coefficient profiles were generated based on the sequence of log (Lambda). When using the value selected by 10-fold cross-validation, the optimal Lambda resulted in 18 non-zero coefficients.

Table 3. Values of various coefficients in the nomogram

Variables	Coef	95%CI
T1.W.M T1-wavelet-LHL_firstorder_Median	-0.5717334	(-2.85 -0.25)
T1.W.HSAHGE T1-wavelet-HLH_glszm_SmallAreaHighGrayLevelEmphasis	0.2269468	(-0.21 - 2.91)
T1.W.CT T1-wavelet-HHL_glcm_ClusterTendency	-2.1615966	(-3.72 - 0.87)
T1.W.R T1-wavelet-HHH_firstorder_InterquartileRange	0.4915525	(-1.02 - 3.32)
T2.O.DE T2-original_gldm_DependenceEntropy	-4.1807337	(-4.97 - 0.68)
T2.W.gl T2-wavelet-HLH_glcm_Idn	47.4461351	(-0.67 - 3.38)
B7.H3	3.6937888	(0.18 - 4.34)
CT	1.4626247	(-1.98 - 4.90)

Table 3: The coefficient values of individual independent factors in the developed column chart are shown in the figure. The selected radiomic features and their abbreviations in the nomogram are presented in the table. All predictive factors have a p-value < 0.05, indicating statistical significance.



3.5 Clinical use

Using the pathological examination results as the gold standard, we calculated the diagnostic sensitivity and specificity of LN status from CT reports, B7-H3 mRNA expression, MRI radiomic features, and the combined predictive model for lymph node metastasis in esophageal cancer. The results are shown in Table 4. Furthermore, we plotted ROC curves (Figure 4) to illustrate the diagnostic performance of LN status from CT reports, B7-H3 mRNA expression, MRI radiomic features, and the established combined predictive model for preoperative lymph node diagnosis. Through comparison, we found that the combined predictive model exhibited the best discriminative power and predictive stability, with the highest AUC value (Table 4, Figure 4).

The decision curve analysis (DCA) based on the combined predictive model is presented in Figure 5. Compared to DCA using a single radiomic feature, the combined predictive model incorporating B7-H3 mRNA expression and clinical CT results demonstrates higher accuracy in predicting preoperative lymph node status. This indicates that the nomogram based on this predictive model is a reliable clinical tool for predicting lymph node status in patients with esophageal cancer. The DCA suggests that within the probability threshold range of approximately 0.3 to 0.7, the nomogram based on the combined model provides additional net benefit to the "treatment" strategy.

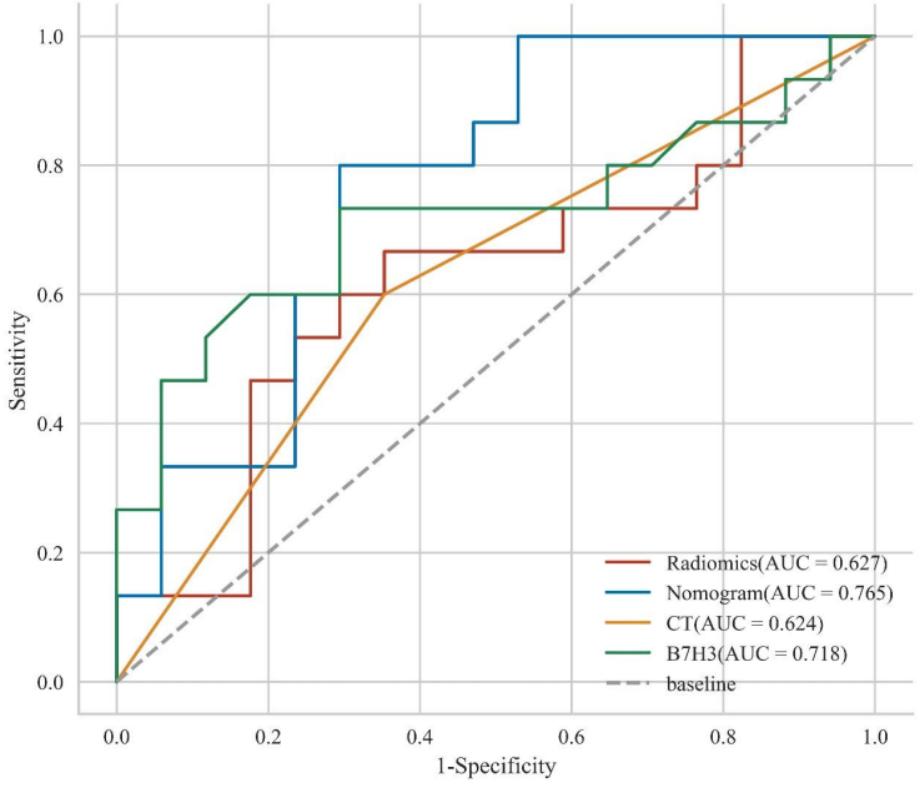


Figure 4: Model performance in the study cohort. The AUC value increases when radiomic features are combined with B7-H3 mRNA expression compared to different feature combinations. Additionally, it can be observed that as more feature species are used, the AUC value increases, indicating better model performance.

Table 4 ROC curve of the dataset

Model	AUC (95%CI)	sensitivity	specificity	Youden index
Radiomics	0.627 (0.427-0.828)	0.667	0.647	0.314
Nomogram	0.765 (0.598-0.931)	0.800	0.706	0.506
CT	0.624 (0.426-0.821)	0.600	0.647	0.247
B7-H3mRNA	0.718 (0.528-0.907)	0.733	0.706	0.439

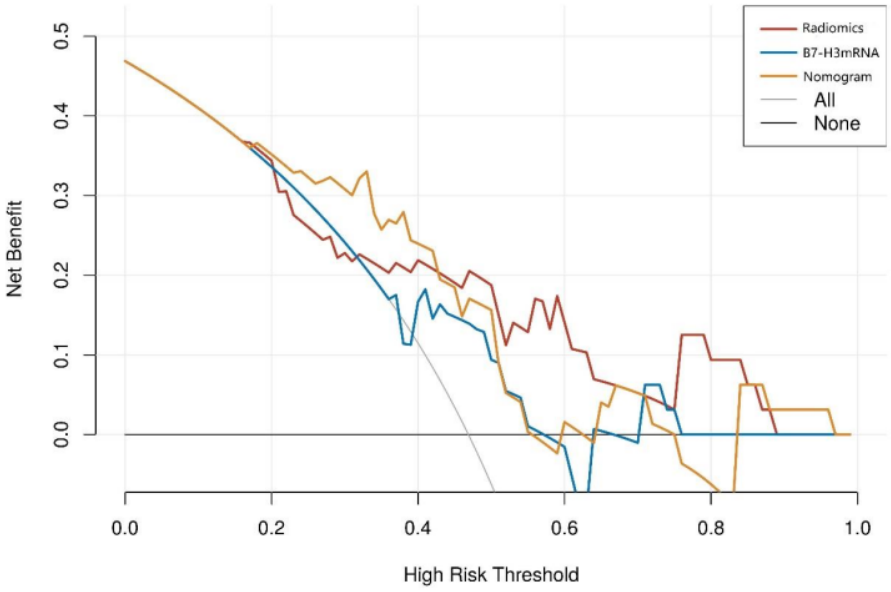


Figure 5: Decision curve analysis (DCA). The y-axis represents net benefit. The threshold probability refers to the point at which the perceived benefit of treating patients with intermediate to high-risk lymph node metastasis (LNM) is considered equivalent to the harm of overtreating low-risk disease, reflecting how patients weigh the benefits and harms associated with decision-making. The higher curve at any given threshold probability represents the optimal prediction that maximizes net benefit. The decision curve indicates that the combined predictive model used provides greater net benefit compared to other models.

4 Discussion

Lymph node status is one of the most important prognostic factors in esophageal cancer, especially the number and location of metastatic lymph nodes, which are closely linked to clinical treatment decisions, including the implementation of neoadjuvant therapy, the extent of surgical lymph node resection, or the design of radiation therapy fields(25, 31, 32). The decision of whether neoadjuvant chemoradiotherapy is required before surgery primarily depends on the lymph node status. Furthermore, neoadjuvant chemoradiotherapy can target micrometastases, including lymph node metastases, and patients with lymph node metastasis may benefit from this treatment(7, 33). However, for patients who refuse or are unable to tolerate surgery, the lymph node status cannot be diagnosed through postoperative biopsy. Therefore, accurate preoperative prediction of lymph node status is necessary and important.

With the advancement of radiomics research, an increasing number of studies are utilizing radiological features extracted from medical images, such as shape, texture, or waveform, to obtain a range of information about cancer phenotypes and the tumor microenvironment(27). This information is distinct and complementary to other relevant data, including clinical features, treatment-related decision information, or genomic data(34). When radiomics-derived data is combined with other relevant data and correlated with clinical disease outcomes, they can generate accurate and reliable Clinical Decision Support Systems (CDSS)(35, 36). These CDSS can assist clinicians in making more informed decisions regarding the diagnosis, treatment planning, and prognosis of cancer patients.

Unlike many previous studies on radiological features, which focused solely on the association of clinical and radiological features with tumor microenvironment characteristics in lymph node metastasis, they neglected the impact of various proteins or tumor factors on promoting lymph node metastasis(37-42). In the study by Toiyama et al., they observed improved predictive accuracy when adding serum biomarkers to the predictive model as clinical pathological risk factors for preoperative detection of lymph node metastasis in colorectal cancer patients (the area under the curve increased to 0.801 [95% CI, 0.725-0.857] with modification of the multivariate model)(43). Similarly, Huang et al. provided a radiomics nomogram incorporating radiomic features, LN status from CT reports, and CEA levels, which demonstrated higher accuracy in the preoperative individualized prediction of lymph node metastasis in colorectal cancer patients(29). These studies may support the notion that considering tumor diagnostic biomarkers across different aspects is an important research approach to enhance CDSS(44).

In our previous studies, we demonstrated a close association between high expression of B7-H3 and tumor differentiation, TNM staging, and lymph node metastasis in esophageal cancer. In this study, we further investigated the correlation between B7-H3 mRNA expression levels and lymph node metastasis in esophageal cancer using the RT-PCR method. Additionally, we developed and validated a diagnostic nomogram based on radiomic features for individualized preoperative prediction of lymph node metastasis in patients with esophageal cancer. The nomogram incorporates three components: radiomic features, B7-H3 mRNA expression levels, and lymph node status from CT reports.

In this study, we determined the expression levels of B7-H3 mRNA in tumor tissues of esophageal cancer using preoperative endoscopic biopsy. We found that the expression levels of B7-H3 mRNA were consistent with previous reports regarding lymph node metastasis in esophageal cancer. Additionally, our research revealed that the relative expression level of B7-H3 mRNA was significantly higher in lymph node-positive tissues compared to lymph node-negative tissues in esophageal squamous cell carcinoma. We performed ROC curve analysis and found that B7-H3 mRNA had good accuracy in predicting lymph node metastasis (AUC=0.718; 95% CI 0.528-0.907; sensitivity: 70.3%; specificity: 70.6%). Some studies on diagnostic models have shown that relying solely on certain factors with univariate associations may not provide sufficient predictive strength(45). However, if a factor has statistical significance, it should not be excluded from the model(46). Therefore, in our study, we found a close association between B7-H3 and lymph node metastasis, indicating the importance of B7-H3 mRNA expression in preoperative lymph node diagnosis.

We performed MRI image acquisition in esophageal cancer patients using T2-TSE-BLADE and T1-StarVIBE sequences, which provide high image quality and anatomical details in EC and accurately depict different layers of the esophageal wall. Therefore, both T2-TSE-BLADE and T1-StarVIBE sequences are feasible for texture analysis(41). In Python, we used the XGBoost model from the Pyradiomics package and employed leave-one-out cross-validation to extract radiomic features from the ROI in MRI. Subsequently, logistic regression analysis was conducted in R language, and based on the AIC criterion, we selected the model with the lowest entropy value. We chose six radiomic features from the extracted features, which primarily represented the tumor's texture complexity and were highly correlated with tumor heterogeneity and prognosis(47).

These extracted radiomic features, along with the B7-H3 mRNA expression and lymph node status from CT reports, were considered as three independent predictive factors. Subsequently, we constructed a combined prediction model using these independent risk factors. Through ROC curve analysis, our constructed combined prediction model demonstrated the highest diagnostic value in predicting lymph node metastasis (AUC=0.765; 95% CI 0.598-0.931; sensitivity: 80.0%; specificity: 70.6%).

We applied DCA to evaluate whether the nomogram of the combined model would improve patient outcomes and thus demonstrate the clinical utility of the nomogram. DCA curves provide insights into clinical consequences based on threshold probabilities and can yield net benefits (defined as the proportion of true positives minus the proportion of false positives)(48). Analysis of the DCA curves revealed that the combined model had higher net benefits at most threshold probabilities, suggesting that the model could be an effective approach to guide clinical decision-making and provide an accurate and reliable CDSS.

For ease of clinical application, based on the three independent risk factors from the aforementioned study, we constructed a clinical radiomic nomogram combining radiomic features, B7-H3 mRNA expression, and lymph node status from CT reports. This nomogram scoring

system can generate the preoperative probability of lymph node metastasis, enabling individualized preoperative prediction of lymph node metastasis risk. Both physicians and patients can utilize the nomogram to make personalized preoperative predictions of lymph node metastasis risk, aligning with the current trend of personalized medicine(49).

Certainly, this study has some limitations. Firstly, the MRI image acquisition did not include the DWI sequence, which could enrich the extracted radiomic feature library and potentially reveal more valuable radiomic features. Although DWI has demonstrated strong capabilities in distinguishing benign and malignant lymph nodes in certain cancers, the respiratory motion specific to the chest can introduce significant artifacts and affect image quality in 3.0T DWI imaging(50). Secondly, our study only used qRT-PCR to demonstrate the mRNA expression level of B7-H3 and did not further investigate corresponding genomic features to analyze the differences between genomic and radiomic features(51). Thirdly, although the predictive model designed in this study showed good accuracy, our sample size was relatively small, and a larger sample size would improve the confidence and performance of the lymph node metastasis prediction model in EC. Fourthly, we analyzed the regions of interest mainly focusing on the primary tumor and did not obtain information from the surrounding tumor-free regions, which may also contain important information. Further research is needed to address these issues.

In conclusion, a correlation was observed between B7-H3 mRNA expression levels and lymph node metastasis in esophageal cancer patients based on preoperative gastric endoscopic specimens. Moreover, a clinical radiomic nomogram incorporating radiomic features, B7-H3 mRNA expression levels, and lymph node status from CT reports was developed, enabling convenient identification of esophageal cancer patients with lymph node metastasis. This nomogram facilitates individualized preoperative prediction of lymph node metastasis in EC patients, thereby providing guidance for the formulation of clinical treatment decisions and facilitating the selection of more rational and effective therapeutic strategies to prevent adverse patient outcomes.

5 Conclusion

This study developed a radiomics nomogram that includes radiomics features, LN status from CT reports, and B7-H3 mRNA expression, enabling convenient preoperative individualized prediction of lymph node metastasis in EC patients.

Article Highlights

The study designed a nomogram that can be applied in a clinical setting to assess the status of preoperative lymph nodes in esophageal cancer patients, aiming to provide better guidance for treatment planning.

89975_Auto_Edited.docx

ORIGINALITY REPORT

9%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.wjgnet.com Internet	149 words — 2%
2	Menglei Li, Jing Zhang, Yibo Dan, Yefeng Yao, Weixing Dai, Guoxiang Cai, Guang Yang, Tong Tong. "A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer", Journal of Translational Medicine, 2020 Crossref	125 words — 2%
3	ascopubs.org Internet	108 words — 2%
4	link.springer.com Internet	84 words — 1%
5	doctorpenguin.com Internet	37 words — 1%
6	aacrjournals.org Internet	36 words — 1%
7	www.frontiersin.org Internet	33 words — 1%

EXCLUDE QUOTES ON EXCLUDE SOURCES < 1%