Volume 21, No 2 | International Journal of Radiation Research, April 2023

Development and validation of a stacking nomogram for predicting regional lymph node metastasis status in rectal cancer via deep learning and hand-crafted radiomics

J. Liu, L. Sun*, X. Lu, Y. Geng, Z. Zhang

Department of Radiology, The fourth Affiliated Hospital of China Medical University, East Chongshan Road, Shenyang, 110000, Liaoning, China

► ***Original article***

Corresponding author:
Lingling Sun, Ph.D.,
E-mail:
liujiaxuan666666@163.com

Received: August 2022
Final revised: December 2022
Accepted: January 2023

Int. J. Radiat. Res., April 2023; 21(2): 267-274

DOI: 10.52547/ijrr.21.2.13

Keywords: *Rectal cancer, lymph node metastasis, radiomics, deep learning, machine learning.*

ABSTRACT

***Background*:** Preoperative assessment of lymph node metastasis (LNM) status is the basis of individual treatment for rectal cancer (RC). However, conventional imaging methods are not accurate enough. ***Materials and Methods*:** We collected 282 RC patients who were divided into the training dataset (n=225) and the test dataset (n=57) with an 8:2 scale. A large number of deep learning (DL) features and hand-crafted radiomics (HCR) features of primary tumors were extracted from the arterial and venous phases of the computed tomography (CT) images. Three machine learning models, including support vector machine (SVM), k-nearest neighbor (KNN),and multi-layer perceptron (MLP) were utilized to predict LNM status in RC patients. A stacking nomogram was constructed by selecting optimal machine learning models for arterial and venous phases, respectively, combined with predictive clinical features. ***Results*:** The stacking nomogram performed well in predicting LNM status, with an area under the curve (AUC) of 0.914 [95% confidence interval (CI): 0.874-0.953] in the training dataset, and an AUC of 0.942 (95%CI: 0.886-0.997) in the test dataset. The AUC of the stacking nomogram were higher than those of CT_reported_N_status, ASVM, and VSVM model in the training dataset (P <0.05). However, in the test dataset, although the AUC of the stacking nomogram was higher than the VSVM, the difference was not obvious (P =0.1424). ***Conclusion*:** The developed deep learning radiomics stacking nomogram showed to be effective in predicting the preoperative LNM status in RC patients.

INTRODUCTION

Globally, colorectal cancer (CRC) is a severe gastrointestinal malignancy, with rectal cancer (RC) accounting for about one-third [1-2]. For RC patients, treatment strategies are determined by risk stratification based on tumor-node-metastasis (TNM) stage [3]. Several studies have demonstrated that preoperative lymph node metastasis (LNM) status is one of the key factors that not only determines the scope of the surgical procedure, but also indicates the therapeutic efficacy [4-8]. Furthermore, it also has a prognostic influence on long-term survival outcomes of RC patients.

Traditional imaging modalities, such as ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), which can provide anatomical and morphological data, are common methods for the clinical diagnosis of RC patients. The American Society of Colon and Rectal Surgeons (ASCRS) demonstrated that the sensitivity and specificity required for detecting LNM status are 55% and 74% by CT, 67% and 78% by endorectal US (EUS), and 66% and 76% by MRI [6], indicating that each of them is not accurate enough for the determination of preoperative LNM status. Additionally, diagnosis through imaging mainly relies on direct observations, reflecting that the diagnostic accuracy may be affected by the radiologist's knowledge, diagnostic experience, employment position, etc. [9-13]. Thus, there is a great clinical need to develop a further accurate method to assess the preoperative LNM status in RC patients.

Radiomics, excavating of invisible features from medical imaging, has noticeably attracted oncologists' attention. Using selected radiomics features, namely handcrafted features, radiomics incorporates a series of computational techniques, including machine learning (ML) algorithms to analyze data to improve diagnostic, prognostic, and predictive accuracy [14-16]. Convolutional neural networks (CNN) are deep learning models, learning from the data itself, whose core layer is convolution, possessing some advantages in terms of dealing with large datasets and the ability to classify and predict the outputs of data analysis [17-19]. Both radiomics and deep learning have shown promising results on CT imaging for determination of LNM status [20, 21]. Stacking is an

[DOI: 10.52547/ijrr.21.2.13]

ensemble learning technique that uses predictions for multiple nodes to construct a new model. The stacking method, which is composed of basic learners and meta-learner, extracts the predictive probability from a ML model (basic learner) as the input variable of the meta-learner to provide a robust model for prediction and classification. Several studies have demonstrated that the performance of stacking method is more stable than that of the individual classifiers [22, 23].

As CT examination is cost-effective, fast, prevalent, and is broadly utilized preoperatively for RC patients [24]. In our study, *we aimed to* develop and validate a CT-based deep learning & radiomics stacking nomogram to predict preoperative LNM status in RC patients, which is rarely seen in our field, which is also the innovation of our study.

MATERIALS AND METHODS

Patients

The acquisition of the imaging data and clinical data of the cases was approved by the institutional ethics committee of our hospital (Ethical review number：EC-2022-KS-035). Among 454 RC patients who were admitted to the Fourth Affiliated Hospital of China Medical University (Shenyang, China) from March 2016 to December 2021, 282 patients were selected (117 female vs. 165 male ; mean age, 65.94±10.76 years old, range of age, 24-91 years old). The inclusion and exclusion criteria are presented in figure 1.

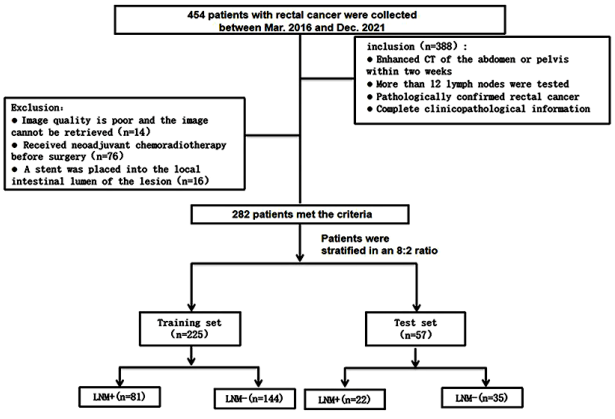

Figure 1. Flowchart of patient recruitment.

Patients were assigned into training dataset (n=225) and test dataset (n=57) by an 8:2 ratio. Baseline clinical characteristics of each patient, including age, gender, smoking history, drinking history, carbohydrate antigen 19-9 (CA19-9) level, cancer antigen-125 (CA125) level, carcinoembryonic antigen (CEA) level, blood routine, and four indicators of blood lipid were obtained from electronic medical history. Two imaging physicians (with 9 and 25 *years' experience* in abdominal imaging) who were blinded to pathologic data of all patients predicted the LNM status of RC patients using CT images, addressed the differences through discussion. The criteria of lymph node metastasis should meet any of the following conditions: (1) uneven reinforcement; (2) irregular boundaries; (3) short diameter>10 mm; (4) 3 or more clustered lymph nodes in the lymph reflux area.

CT imaging acquisition

All patients underwent biphasic (arterial and venous phases) enhanced CT scan using Philips iCT 256 spiral CT scanner (Philips, Amsterdam, Netherlands) preoperatively. CT parameters are as follows: tube voltage 120 kV, tube current automatic regulation, pitch 0.5s, transverse fault thickness 5mm, layer spacing 5mm, and matrix 512×512. The patient was taken in the supine position and injected with the contrast ioxol (300 mg/mL) at 80 to 100 mL, flow rate of 3.0-3.5mL/s with a delay of about 30 to 35s and 60 to 70s, to obtain enhanced abdominal CT images during the arterial and venous phases, respectively.

Feature extraction and feature screening

We imported all the CT images into the open-source 3d-slicer software (www.3D-Slicer.com, version 4.13.2). Firstly, images were converted into a standardized input with an intensity range of -1024 to +1024 HU using a uniform abdominal window (window level [WL]=50 and window width [WW] =350). Then, the two radiologists manually contour and segmented the primary tumors from the axial CT images at the arterial and venous phase. The regions of interest (ROI) includes areas of necrosis or bleeding, but avoids normal large bowel walls and bowel contents as much as possible. Finally, two ROI (arterial and venous phase primary tumors) were generated for each patient.

Subsequently, we resampled CT images to voxel sizes of 3.0mm.The original images were preprocessed by high and low wavelet filters, and using the square, square-root, logarithmic, exponential, and gradient transformations. Hand-crafted features were extracted using the PyRadiomics *(Python 3.7.1, version: 3.0.1)*. Hand-crafted features included first-order (n=252), including gray-level cooccurrence matrix (GLCM, n=308), gray-level difference matrix (GLDM, n=196), gray-level size zone matrix (GLSZM, n=224), and grey -level run-length matrix (GLRLM, n=224).

In addition, a CNN ResNet50 model was used for deep learning feature extraction. First, the maximum cross section of lesions was selected from the manually delineated ROIs of arterial phase and venous phase as the input model, the features of the avgpool layer of the model were extracted, and the deep learning features of arterial phase and venous phase were obtained.

A total of 3254 features (2048 deep learning

[DOI: 10.52547/ijrr.21.2.13]

features and 1204 hand-crafted features) were extracted from the segmented ROI of each CT image. Normalization of all features was carried out to a standardized numerical range. First, the intraclass correlation coefficient (intra-ICC)/interclass correlation coefficient (inter-ICC) ratio was used to assess the stability of two radiologists' results in the tumor delineation. Features with an ICC> 0.75 were retained. The correlation coefficients within the features were calculated using the Spearman correlation with a threshold of 0.9, retaining only one feature when it was highly correlated. Then, the Mann-Whitney U test (P<0.05) was used to eliminate redundant features that were ineffective for classification. Finally, the lasso regression combined with cross-validation was utilized to screen the most predictive features from the remaining features. The study workflow is shown in figure 2.

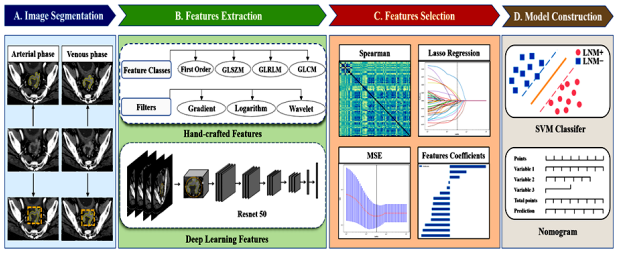

Figure 2. Workflow for this deep learning and radiomics study, consisting of image segmentation, feature extraction, feature selection, and model construction.

Construction and evaluation of the predictive model

Three ML models, including SVM, MLP, and KNN were used to develop prediction models (at arterial and venous phases) for LNM status before surgery, and the test dataset was utilized to examine the effectiveness of the ML models derived from the training set.

We calculated the AUC of the receiver operating curve (ROC) and the ACC of three ML models in the training dataset and test dataset to calculate their 95% CI. The three ML models were compared, and the most efficient ML model was selected. The most efficient ML models for the arterial and venous phases were classified as Amodel and Vmodel, respectively. In addition, predictive clinical features were selected using Fisher's exact test, Chi-square test and the t-test (P<0.05). Finally, a stacking nomogram was constructed via combining the most efficient ML models of the arterial and venous phases and the predictive clinical features. This stacking nomogram was validated by the Hosmer-Lemeshow test.

The performance of the clinical models, Amodel, Vmodel, and stacking nomogram was evaluated using accuracy, AUC, specificity, and sensitivity .ROC curves were compared using the DeLong test. We chose calibration curves to measure the consistency of the predicted and true values of the stacking nomogram. The decision curve is used to evaluate the standardized net benefit at different thresholds for different four models.

Statistical analysis

We used the R software (https:www.r-project.org, version:4.1.2) and Anaconda (https:www.anaconda.com, version: Python3.7) *for t*he statistical analysis. We used independence t-test to compare continuous type variables and we used the Chi-square test or Fisher's exact test for comparison of categorical variables. P<0.05 of two-sided was seen statistically significant.

RESULTS

Clinical characteristics

The study included a total of 282 patients (65.94 ± 10.76 years old; mean age, 65.94 years old). CT_reported_N_status was statistically different between the LNM- and LNM+ groups (P<0.05). There were no significant differences (P>0.05) between the LNM- and LNM+ groups in terms of age, gender, smoking history, CEA level, CA12-5 level, CA19-9 level, blood routine index, and four indicators of blood lipid in the training and test datasets. Patients' basic clinical characteristics are presented in table 1.

Feature screening

A total of 3254 features (2048 deep learning features and 1204 hand-crafted features) were extracted from the ROI of each CT segment. Firstly, the total arterial / venous phase features have been excluded for 83 and 94 features, respectively, according to ICC<0.75. Additionally, 2249/2252 features were retained based on Spearman correlation coefficient analysis. Then, 64 and 57 features were screened out through Mann-Whitney U test. Ultimately, the final 16 and 14 arterial / venous features were determined by Lasso regression. The λ value of the minimum error was chosen as the parameter to determine 16 arterial phase (lambda = 0.0395) and 14 venous phase (lambda=0.0327) features. The selected features and the feature weights are shown in figures 3A & 3B.

Development of the ML models

The AUC values and ROC curves of the three ML models are shown in table 2 and figure 4**.** The AUC values of the three machine models indicate that the established radiomics model can predict the preoperative RC LNM status and have a satisfactory performance. According to the results of Delong test, we selected SVM as the best model of arterial phase and venous phase, and the accuracy of the arterial phase SVM (ASVM) and venous phase SVM (VSVM) training sets was 0.787 and 0.791, specificity of 0.743 and 0.826, and sensitivity of 0.864 and 0.728, respectively. The arterial phase (ASVM) and venous

[DOI: 10.52547/ijrr.21.2.13]

phase (VSVM) test sets had accuracy of 0.754 and 0.842, specificity of 0.686 and 0.800, and sensitivity of 0.864 and 0.909 respectively. In the training dataset, the AUC of SVM was higher than the KNN and MLP in both arterial and venous phases (P<0.05). However, in the test dataset, although the AUC of ASVM was higher than the AMLP, the difference was not notable (P=0.4234).

Table 1. Statistical analysis results of clinical characteristics

Characteristic	Training set (n=225)		P	Test set (n=57)		P
	LNM-(n=144)	LNM+(n=81)		LNM-(n=35)	LNM+(n=22)	
Basic clinical data						
Age, (mean ± SD) (years)	66.99±10.37	64.51±11.56	0.100	64.83±10.92	66.09±9.78	0.660
Gender			0.202			0.072
Male	89(84.5)	43(47.5)		17(20.3)	17(20.3)	
Female	55(59.5)	38(33.5)		18(14.7)	6(9.3)	
Smoking			0.749			0.682
No	102(103.0)	59(58.0)		22(22.7)	15(14.3)	
Yes	42 (41.0)	22(23.0)		13(12.3)	7(7.7)	
Routine blood test						
HGB (g/L)	129.46±19.17	125.35±22.95	0.152	132.17±13.18	123.32±21.31	0.057
RBC (10^{12}/L)	4.37±0.58	4.24±0.55	0.104	4.35±0.43	4.20±0.38	0.190
WBC (10^9/L)	6.47±1.85	6.79±2.12	0.237	6.55±1.96	6.54±1.83	0.986
PLT (10^9/L)	229.32±71.60	242.03±83.10	0.229	224.66±78.38	244.12±76.23	0.360
Lymphocyte(10^9/L)	1.58±0.58	1.56±0.55	0.792	1.65±0.62	1.95±1.27	0.231
Monocyte(10^9/L)	0.46±0.23	0.46±0.18	0.927	0.42±0.15	0.66±1.00	0.166
Neutrophil(10^9/L)	4.25±1.61	4.55±1.95	0.215	4.24±1.93	4.22±0.97	0.973
Lipid metabolism in serum (mmol/L)						
TG	1.49±1.15	1.33±0.55	0.244	1.43±0.86	1.40±0.60	0.882
Cholesterol	4.77±0.92	4.64±0.82	0.276	4.57±1.13	4.62±1.16	0.877
HDL	1.23±0.83	1.17±0.29	0.516	1.16±0.32	1.08±0.27	0.338
LDL	2.94±0.72	2.85±0.75	0.376	2.76±0.97	3.03±1.16	0.347
Serum tumor markers						
CEA (≥5ng/mL)			0.526			**0.012***
No	88(85.8)	46(48.2)		26(21.5)	9(13.5)	
Yes	56(58.2)	35(32.8)		9(13.5)	13(8.5)	
CA19-9 (≥37U/mL)			0.289			0.075
No	128(125.4)	68(70.6)		32(29.5)	16(18.5)	
Yes	16(18.6)	13(10.4)		3(5.5）	6(3.5)	
CA12-5 (≥30U/mL)			0.620			0.053
No	142(141.4)	79(79.6)		35(33.2)	19(20.8)	
Yes	2(2.6)	2(1.4)		0(1.8)	3(1.2)	
CT reported N status			**0.000***			**0.001***
0	101(83.2)	29(46.8)		22(16.0)	4(10.0)	
1	43(60.8)	52(34.2)		13(19.0)	18(12.0)	

SD, standard deviation; HDL,high density lipoprotein ;LDL,low density lipoprotein ;TG,triglyceride;
CEA, carcinoembryonic antigen; CA 19–9, carbohydrate antigen 19–9; CA 125, carbohydrate antigen 125 , *P < 0.05

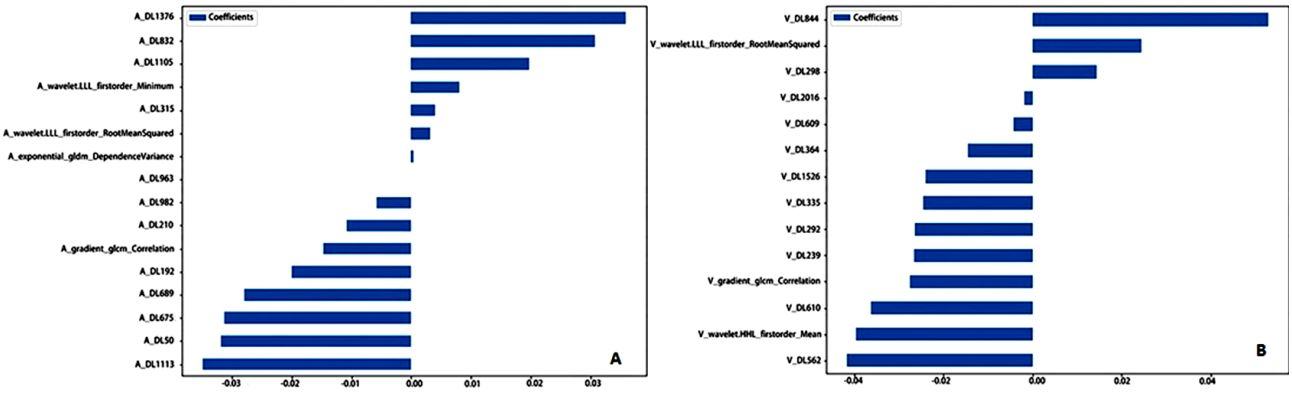

Figure 3. 16 selected arterial phase features **(A)** and 14 selected venous phase features **(B)**.

[DOI: 10.52547/ijrr.21.2.13]

Table 2. Arterial and Venous phase prediction performance of three machine models

model-name	Accuracy	AUC	95%CI	Sensitivity	Specificity	dataset
ASVM	0.787	0.849	0.796~0.902	0.864	0.743	training set
ASVM	0.754	0.851	0.751~0.951	0.864	0.686	testing set
AKNN	0.738	0.774	0.714~0.833	0.679	0.771	training set
AKNN	0.614	0.742	0.617~0.867	0.909	0.429	testing set
AMLP	0.698	0.795	0.737~0.854	0.889	0.597	training set
AMLP	0.825	0.805	0.676~0.934	0.591	0.971	testing set
VSVM	0.791	0.851	0.800~0.903	0.728	0.826	training set
VSVM	0.842	0.892	0.808~0.977	0.909	0.800	testing set
VKNN	0.729	0.758	0.695~0.821	0.605	0.799	training set
VKNN	0.649	0.770	0.651~0.889	0.909	0.486	testing set
VMLP	0.729	0.790	0.730~0.851	0.840	0.667	training set
VMLP	0.737	0.764	0.639~0.889	0.818	0.686	testing set

ASVM:Arterial phase support vector machine； VSVM:Venous phase support vector machine;
AKNN:Arterial phase k-nearest neighbor;VKNN:Venous phase k-nearest neighbor;
AMLP:Arterial phase multi-layer perceptron; VMLP:Venous phase multi-layer perceptron;

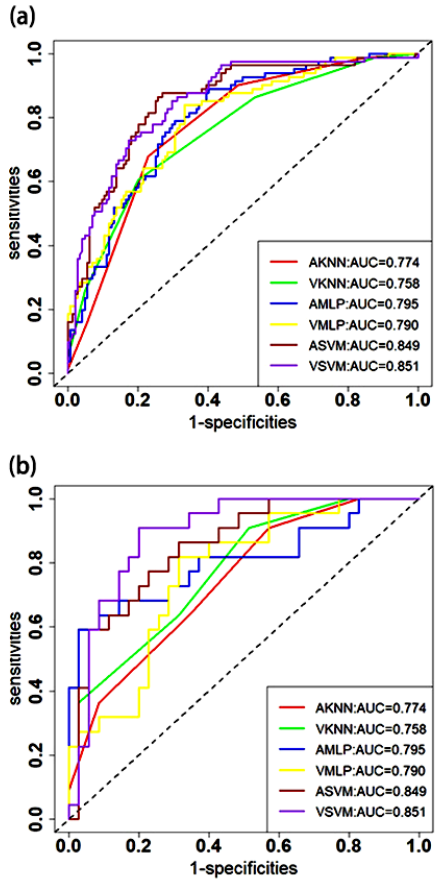

Figure 4. Receiver operating characteristic (ROC) curves at arterial phase and venous phase in the training set **(a)** and test set **(b)** for three machine models: KNN, MLP, SVM.

Construction of stacking nomogram

We recorded the probability of predicting LNM+ by the SVM model in the arterial and venous phase as ASVMP and VSVMP, respectively. The ASVMP and VSVMP were integrated with CT_reported_N_status to construct a stacking nomogram by the logistic regression, as shown in figure 5. The Hosmer-Lemeshow test indicated that the stacking nomogram has a good fit (P=0.149). The performance of the CT_reported_N_status, ASVM, VSVM, and stacking nomogram was assessed and compared (table 3). The AUC of the stacking nomogram in the training dataset (stacking nomogram vs. ASVM vs. VSVM vs. CT_reported_N_status, 0.914 vs. 0.849 vs. 0.851 vs. 0.672) and the test dataset (stacking nomogram vs. ASVM vs. VSVM vs. CT_reported_N_status, 0.942 vs. 0.851 vs. 0.892 vs. 0.723) was superior to ASVM, VSVM, and CT_reported_status. However, the DeLong test showed that the ROC curves of the stacking nomogram and the VSVM did not significantly differ in the test dataset (P=0.1424). Figure 6 shows the calibration curve of the stacking nomogram, indicating that the predicted values are in good agreement with the actual values. The decision curve (figure 7) illustrates a good net benefit of LNM. Compared with the other three models (CT_reported_N_status, ASVM, and VSVM), RC patients could benefit more from the stacking nomogram when the probability thresholds in the training dataset and the test dataset are between 0.2 and 0.7.

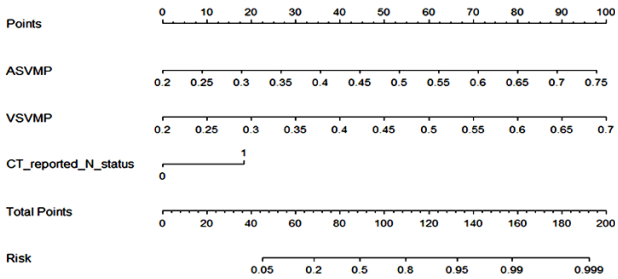

Figure 5. Stacking nomogram combining ASVMP, VSVMP, and CT_reported_N_status.

Table 3. Prediction performance of four models in the training and test set and test set.

model-name	Accuracy	AUC	95%CI	Sensi-tivity	speci-ficity	dataset
CT_reported_status	0.680	0.672	0.607~0.736	0.642	0.701	training set
CT_reported_status	0.702	0.723	0.608~0.839	0.818	0.629	test set
ASVM	0.787	0.849	0.796~0.902	0.864	0.743	training set
ASVM	0.754	0.851	0.751~0.951	0.864	0.686	test set
VSVM	0.791	0.851	0.800~0.903	0.728	0.826	training set
VSVM	0.842	0.892	0.808~0.977	0.909	0.800	test set
Stacking nomogram	0.858	0.914	0.874~0.953	0.901	0.833	training set
Stacking nomogram	0.895	0.942	0.886~0.997	0.955	0.857	test set

[DOI: 10.52547/ijrr.21.2.13]

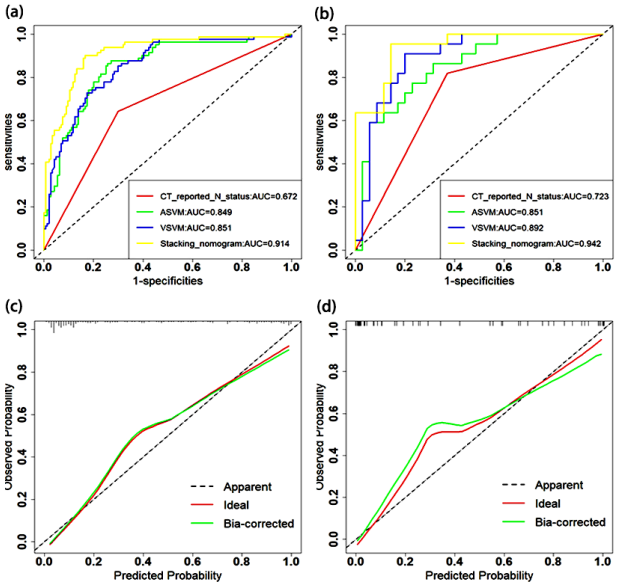

Figure 6. Receiver operating characteristic (ROC) curves for training **(a)** and test set **(b)** for four models: CT_reported_N_status, ASVM, VSVM and stacking nomogram. Stacking nomogram calibration curves for each set. **(c)** The calibration curves of the stacking nomograms in the training set. **(d)** The calibration curves of the stacking nomograms in the test set. The X-axis represents the predicted risk of lymph node metastasis. The Y-axis represents the actual lymph node metastasis rate.

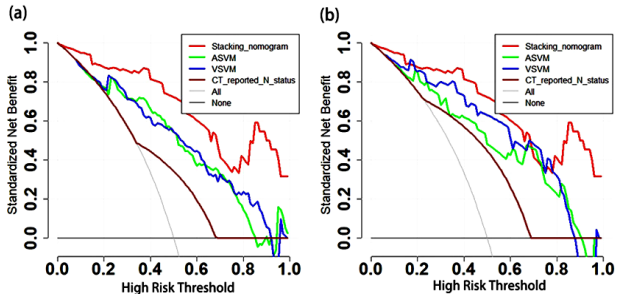

Figure 7. Decision curve analysis of the CT_reported_N_status, ASVM, VSVM and stacking nomogram in the training set **(a)** and test set **(b)**. The Y-axis measures the net benefit. The brown, blue, green, and red lines represent CT-reported-N-status, ASVM, VSVM and Stacking nomogram, respectively. The black line indicates the hypothesis that no patients would develop lymph node metastasis, and the grey line indicates the hypothesis that all patients would develop lymph node metastasis. The result of the curve shows that the stacking nomogram has better predictive ability than ASVM (green), VSVM (blue), and CT-reported-N-status (brown) when the high-risk threshold (X-axis) is 0.2-0.7 both training set **(a)** and test set **(b)**.

DISCUSSION

It is noteworthy that LNM status is a key determinant of neoadjuvant therapy or surgical resection. The present study explored the value of deep learning radiomics based on biphasic contrast-enhanced CT in the preoperative prediction of LNM status in RC patients. Previous studies have mainly evaluated the characteristics of single-phase CT. The present study provided a promising prediction model, according to the values of AUC, sensitivity, and specificity, for the LNM status in RC patients based on the CT imaging features at arterial and venous phases. Non-contrast-enhanced CT radiomics model was not used; on the one hand, the external contour range of the lesion appeared after contrast enhancement for more predictive information; on the other hand, the non-contrast-enhanced CT segmentation of the tumor lesion has some limitations, especially when the early-stage primary tumor is small, which may lead to some unnecessary errors. Several studies [20, 25] have shown that the contrast-enhanced CT radiomics model will be significantly superior to the non-contrast-enhanced CT radiomics model. Cheng *et al.* [25] predicted LNM status in colorectal cancer patients using non-contrast-enhanced CT, arterial phase CT, and venous phase CT with AUC values of 0.636, 0.728, and 0.690, respectively, suggesting that the results of arterial phase CT and venous phase CT were superior to non-contrast-enhanced CT. Therefore, we recommend the use of contrast-enhanced CT to evaluate the LNM status in the next studies.

The ability of radiomics to analyze the entire tumor volume eliminates bias error similar to pathological sampling, which is also a potential advantage of radiomics analysis [26]. In the present study, it was attempted to evaluate 3,254 CT features at arterial and venous phases, including 2,048 deep learning and 1,204 hand-crafted features, and higher-order features were selected for the radiomics analysis, such as exponential, gradient, logarithmic, square, square root, and wavelet features. Hand-crafted features (3 out of 7) were identified as textural features, including gradient_glcm_Correlation and exponential_gldm_Dependence Variance, and the remaining were first-order features, reflecting differences in tumor intensity and intra-tumor texture in images. This indicates that tumor heterogeneity may be related to LNM status, because first-order-, GLCM-, and GLDM-based textural features are mainly identified to reflect intratumoral heterogeneity and the irregularity of the composition [26-28], confirming the superiority of textural features in determining LNM status. Previous studies [29-30] have shown that textural analysis is advantageous for diagnosis, guiding treatment, and predicting prognosis. NaLaeEun *et al.* [29] demonstrated that textural features were associated with a complete pathological response after neoadjuvant chemoradiotherapy of breast cancer patients. Texture features could differentiate KRAS status based on T2-MRI images [30].

In addition, numerous studies on disease classification, differential diagnosis, and predictive prognostic analysis showed that deep learning could

[DOI: 10.52547/ijrr.21.2.13]

better promote the radiomics analysis [31-33]. Most of the features that we screened (23 out of 30) were deep learning features, suggesting that deep learning plays a more pronounced predictive role in regional LNM in RC, which is similar to previously reported findings [21,34-35]. Using the deep learning models constructed by the Resnet-50 algorithm, we could obtain ML models with good feature learning and feature representation capabilities. The ResNet is mainly based on the residual learning mechanism, which is not only simple, fast, efficient, and accurate, but also has a better performance in object detection, image segmentation and classification [36, 37]. Studies have shown that the Resnet50-based CNN algorithms could be used to detect and classify clinicopathological features with satisfactory results [38-39]. A study on the benign and malignant diagnosis of pulmonary nodules showed a diagnostic accuracy of 87.3% and an AUC of 0.907 using the three-dimensional (3D) ResNet50 [39].

A nomogram for predicting the preoperative LNM status in colorectal cancer was proposed by Huang *et al.* [40], including the radiomics signature, CT-reported LN status, and CEA level. The stacking nomogram proposed in the current study, based on two most efficient ML models in the arterial and venous phases, and CT-reported LN status, also showed a strong predictive capability and is higher than the previous studies. Some studies have shown that a high preoperative CEA level in patients may increase the risk of LNM and may require regular follow-up for close monitoring of LNM status in RC patients [41-42]. The preoperative CEA level in the present study was not statistically significant in predicting LNM status in RC patients, which is consistent with the data from previous studies [43, 44], and this could be related to the noticeable heterogeneity in the expression of rectal cancer in multiple molecules, including CEA. It has been shown that about 70% of colorectal tumors are mainly composed of CEA-negative cell lines, with extremely low or no CEA secretion [45]. In addition, elevated circulating CEA level was found in some benign tumors and inflammatory diseases, and the specificity of CEA for the diagnosis of primary gastrointestinal tumors is limited [46, 47].

In the present study, three ML models of MLP, SVM, and KNN were used, and it was attempted to select the most efficient ML model as the first layer input model of the stacking nomogram, which not only improved the final model "fault tolerance", but also the three ML models exhibited a promising capability for prediction of the LNM status in RC patients. Then, a 2-layer stacking ensemble model was established, consisting of base learners and meta -learners. We employed 3 models (ASVM, VSVM, and CT _ report _ status) as the base learners. Each base learner generated a predictive value for a given binary outcome of the RC patients' LNM status. For the meta-learners, 3 predicted values from the base learners were used as input variables for the final prediction. The stacking nomogram had AUCs ranging from 0.914 to 0.942, which is higher than the range reported recently [25, 48]. The nomogram model constructed by Su et al. based on T2WI radiomics has a good diagnostic value for LNM in RC patients, with an AUC value of 0.891-0.902 [48]. A systematic review and meta-analysis showed that the AUC of per-patient was 0.808 (0.739-0.876) and 0.917 (0.882-0.952) in radiomics and deep learning models of RC, respectively [49], which is similar to the results of the present study. This suggests that deep learning radiomics can be used for tumor diagnosis and treatment guidance in clinical practice through non-invasive evaluation of human tissues, thereby making the treatment further appropriate, particularly for "personalized medicine".

The present study has some limitations. First, the limited number of patients were all from our hospital located at the Northeast China. Although the present study showed satisfactory preliminary results, further multicenter study is required. Second, the present study only evaluated the characteristics of the primary tumor without considering the surrounding lymph nodes, which could be due to the challenge of difficulty in comparing the imaging and pathology of the lymph nodes. Last but not least, deep learning in this study only used images of the maximum cross-sectional area of tumors as input to the ResNet-50 model. The use of 3D tumors should be considered in the future deep learning models.

In all, we established a deep learning radiomics stacking nomogram model to predict the status of regional lymph node metastasis before treatment, which shows high predictive ability and clinical utility to assist clinicians in diagnosis and treatment decisions.

ACKNOWLEDGMENTS

*Thanks to t*he fourth Affiliated Hospital of China Medical University *for technical support of this study. The authors thank medsci Medical Services (*medsci.cn/bioon.com*) English language editing and review services.*

***Declaration of competing interests*:** The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of this article.

***Ethical statement*:** This retrospective study was approved by the medical ethics committee of our hospital (Ethical review number：EC-2022-KS-035).

***Funding*:** This research did not receive any specific grant from funding agencies in the public, commercials, or not-for-profit sectors.

***Author's contribution*:** J.L: Methodology, Conceptual-ization, Data curation, Formal analysis, Writing (original draft), and Validation; L.S: Supervision and Resources; X.L: Supervision, Methodology,

[DOI: 10.52547/ijrr.21.2.13]

Investigation, and Conceptualization; Z.Z.: Data curation and Formal analysis; Y.G.: Data curation and Formal analysis

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin,* ***71*** *(3): 209-249.*
2. Siegel RL, Miller KD, Fuchs HE, Jemal A (2021) Cancer statistics, 2021. *CA Cancer J Clin,* ***71****(1): 7-33.*
3. Ponnatapura J and Lalwani N(2021) Imaging of Colorectal Cancer: Screening, Staging, and Surveillance. *Semin Roentgenol,* ***56****(2): 128 -139.*
4. Benson AB, Venook AP, Al-Hawary MM, *et al.* (2020) NCCN Guidelines Insights: Rectal Cancer, Version 6.2020. *J Natl Compr Canc Netw,* ***18****(7): 806-815.*
5. Borgheresi A, De Muzio F, Agostini A, *et al.* (2022) Lymph Nodes Evaluation in Rectal Cancer: Where Do We Stand and Future Perspective. *J Clin Med,* ***11****(9): 2599.*
6. You YN, Hardiman KM, Bafford A, *et al.* (2020) The American Society of Colon and Rectal Surgeons Clinical Practice Guidelines for the Management of Rectal Cancer. *Dis Colon Rectum,* ***63****(9): 1191-1222.*
7. Lambregts DMJ, Bogveradze N, Blomqvist LK, *et al.* (2022) Current controversies in TNM for the radiological staging of rectal cancer and how to deal with them: results of a global online survey and multidisciplinary expert consensus. *Eur Radiol,* ***32****(7): 4991-5003.*
8. Bates DDB, Homsi ME, Chang KJ, *et al.* (2022) MRI for Rectal Cancer: Staging, mrCRM, EMVI, Lymph Node Staging and Post-Treatment Response. *Clin Colorectal Cancer,* ***21****(1): 10-18.*
9. Kijima S, Sasaki T, Nagata K, *et al.* (2014) Preoperative evaluation of colorectal cancer using CT colonography, MRI, and PET/CT. *World J Gastroenterol,* ***20****(45): 16964-75.*
10. Cianci R, Cristel G, Agostini A, *et al.* (2020) MRI for Rectal Cancer Primary Staging and Restaging After Neoadjuvant Chemoradiation Therapy: How to Do It During Daily Clinical Practice. *Eur J Radiol,* ***131:*** *109238.*
11. Zhuang Z, Zhang Y, Wei M, *et al.* (2021) Magnetic resonance imaging evaluation of the accuracy of various lymph node staging criteria in rectal cancer: A systematic review and meta-analysis. *Front Oncol,* ***11:*** *709070.*
12. Fritz S, Killguss H, Schaudt A, *et al.* (2021) Preoperative versus pathological staging of rectal cancer-challenging the indication of neoadjuvant chemoradiotherapy. *Int J Colorectal Dis,* ***36****(1): 191-194.*
13. Reali C, Bocca G, Lindsey I, *et al.* (2022) Influence of incorrect staging of colorectal carcinoma on oncological outcome: are we playing safely? *Updates Surg,* ***74****(2): 591-597.*
14. Lafata KJ, Wang Y, Konkel B, *et al.* (2021) Radiomics: a primer on high-throughput image phenotyping. *Abdom Radiol,* ***47****: 2986–3002.*
15. Attanasio S, Forte SM, Restante G, *et al.* (2020) Artificial intelligence, radiomics and other horizons in body composition assessment. *Quant Imaging Med Surg,* ***10****(8): 1650-1660.*
16. Avanzo M, Wei L, Stancanello J, *et al.*(2020)Machine and deep learning methods for radiomics. *Med Phys,* ***47****(5): e185-e202.*
17. Jiang Y, Yang M, Wang S, *et al.* (2020) Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun (Lond),* ***40****(4): 154-166.*
18. Kim M, Yun J, Cho Y, *et al.* (2019)Deep Learning in Medical Imaging. *Neurospine,* ***16****(4): 657-668.*
19. Wainberg M, Merico D, Delong A, Frey BJ (2018) Deep learning in biomedicine. *Nat Biotechnol,* ***36****(9): 829-838.*
20. Dong D, Fang MJ, Tang L, *et al.* (2020) Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol,* ***31****(7): 912-920.*
21. Li J, Zhou Y, Wang P, *et al.* (2021) Deep transfer learning based on magnetic resonance imaging can improve the diagnosis of lymph node metastasis in patients with rectal cancer. *Quant Imaging Med Surg,* ***11****(6): 2477-2485.*
22. Kim C, You SC, Reps JM, *et al.* (2021) Machine-learning model to predict the cause of death using a stacking ensemble method for observational data. *J Am Med Inform Assoc,* ***28****(6): 1098-1107.*
23. Yang Y, Wei L, Hu Y, *et al.* (2021) Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning. *J Neurosci Methods,* ***350:*** *109019.*
24. Lokuhetty N, Seneviratne SL, Rahman FA, Marapana T, Niloofa R, De Zoysa I. (2021)Radiological staging of rectal cancer in a resource limited setting. *BMC Res Notes.* ***13(1):*** *479.*
25. Cheng Y, Yu Q, Meng W, Jiang W (2022) Clinico-Radiologic Nomogram Using Multiphase CT to Predict Lymph Node Metastasis in Colon Cancer. *Mol Imaging Biol,* ***24(5):****798-806.*
26. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. (2019)Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. *Comput Struct Biotechnol J.* ***17:*** *995-1008.*
27. Gardin I, Grégoire V, Gibon D, Kirisli H, Pasquier D, Thariat J, *et al.* (2019) Radiomics: Principles and radiotherapy applications. *Crit Rev Oncol Hematol.* ***138:*** *44-50.*
28. Litvin AA, Burkin DA, Kropinov AA, Paramzin FN. (2021) Radiomics and Digital Image Texture Analysis in Oncology (Review). *Sovrem Tekhnologii Med.* ***13(2):*** *97-104.*
29. Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim JA, *et al*. (2020) Texture Analysis with 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer. *Radiology.* ***294(1):*** *31-41.*
30. Oh JE, Kim MJ, Lee J, Hur BY, Kim B, Kim DY, *et al*. (2020)Magnetic Resonance-Based Texture Analysis Differentiating KRAS Mutation Status in Rectal Cancer. *Cancer Res Treat.* ***52(1):*** *51-59.*
31. Bo L, Zhang Z, Jiang Z, Yang C, Huang P, Chen T, *et al*. (2021) Differentiation of Brain Abscess From Cystic Glioma Using Conventional MRI Based on Deep Transfer Learning Features and Hand-Crafted Radiomics Features. *Front Med (Lausanne),* ***8:*** *748144.*
32. Gao R, Zhao S, Aishanjiang K, Cai H, Wei T, Zhang Y, *et al.* (2021) Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. *J Hematol Oncol,* ***14(1):*** *154.*
33. Liu H, Yin H, Li J, Dong X, Zheng H, Zhang T, *et al*. (2022) A Deep Learning Model Based on MRI and Clinical Factors Facilitates Noninvasive Evaluation of KRAS Mutation in Rectal Cancer. *J Magn Reson Imaging,* ***56(6):****1659-1668.*
34. Zhao X, Xie P, Wang M, Li W, Pickhardt PJ, Xia W, *et al*. (2020) Deep learning-based fully automated detection and segmentation of lymph nodes on multiparametric-mri for rectal cancer: A multicentre study. *EBioMedicine,* ***56:*** *102780.*
35. Zhao J, Wang H, Zhang Y, Wang R, Liu Q, Li J, *et al.* (2022) Deep learning radiomics model related with genomics phenotypes for lymph node metastasis prediction in colorectal cancer. *Radiother Oncol,* ***167:*** *195-202.*
36. He K, Liu X, Li M, Li X, Yang H, Zhang H (2020) Noninvasive KRAS mutation estimation in colorectal cancer using a deep learning method based on CT imaging. *BMC Med Imaging,* ***20(1):*** *59.*
37. Chen T, Liu S, Li Y, Feng X, Xiong W, Zhao X, *et al. (*2019) Developed and validated a prognostic nomogram for recurrence-free survival after complete surgical resection of local primary gastrointestinal stromal tumors based on deep learning. *EBioMedicine,* ***39:*** *272-279.*
38. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. (2022) Performance of deep convolutional neural network for classification and detection of oral potentially malignant disorders in photographic images. *Int J Oral Maxillofac Surg,* ***51(5):*** *699-704.*
39. Yu H, Li J, Zhang L, Cao Y, Yu X, Sun J (2021) Design of lung nodules segmentation and recognition algorithm based on deep learning. *BMC Bioinformatics,* ***22(Suppl 5):*** *314.*
40. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, *et al.* (2016) Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol,* ***34(18):*** *2157-64.*
41. Li M, Zhang J, Dan Y, Yao Y, Dai W, Cai G, *et al.* (2020) A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer. *J Transl Med,* ***18(1):*** *46.*
42. Yang YS, Feng F, Qiu YJ, Zheng GH, Ge YQ, Wang YT (2021) High-resolution MRI-based radiomics analysis to predict lymph node metastasis and tumor deposits respectively in rectal cancer. *Abdom Radiol (NY),* ***46(3):*** *873-884.*
43. He J, Wang Q, Zhang Y, Wu H, Zhou Y, Zhao S (2021) Preoperative prediction of regional lymph node metastasis of colorectal cancer based on 18F-FDG PET/CT and machine learning. *Ann Nucl Med,* ***35 (5):*** *617-627.*
44. Liu X, Yang Q, Zhang C, Sun J, He K, Xie Y, *et al*. (2021) Multiregional-Based Magnetic Resonance Imaging Radiomics Combined With Clinical Data Improves Efficacy in Predicting Lymph Node Metastasis of Rectal Cancer. *Front Oncol,* ***10:*** *585767.*
45. Shen D, Wang X, Wang H, Xu G, Xie Y, Zhuang Z, *et al.* (2022) Current Surveillance After Treatment is Not Sufficient for Patients With Rectal Cancer With Negative Baseline CEA. *J Natl Compr Canc Netw*, ***20(6):****653.*
46. Huang R, Meng T, Zha Q, Cheng K, Zhou X, Zheng J, *et al.* (2022) The predicting roles of carcinoembryonic antigen and its underlying mechanism in the progression of coronavirus disease 2019. *Crit Care,* ***25(1):*** *234.*
47. Hao C, Zhang G, Zhang L (2019) Serum CEA levels in 49 different types of cancer and noncancer diseases. *Prog Mol Biol Transl Sci,* ***162:*** *213-227.*
48. Su Y, Zhao H, Liu P, Zhang L, Jiao Y, Xu P, Lyu Z, Fu P (2022) A nomogram model based on MRI and radiomic features developed and validated for the evaluation of lymph node metastasis in patients with rectal cancer. *Abdominal Radiology (New York),* ***47(12):*** *4103–4114.*
49. Bedrikovetski S, Dudi-Venkata NN, Kroon HM, Seow W, Vather R, Carneiro G, *et al.* (2021) Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. *BMC Cancer,* ***21(1):*** *1058.*

[DOI: 10.52547/ijrr.21.2.13]