Research article

Predicting rectal cancer tumor budding grading based on MRI and CT with multimodal deep transfer learning: A dual-center study

Ziyan Liu [a,1], Jianye Jia [a,1], Fan Bai [a], Yuxin Ding [a], Lei Han [b], Genji Bai [a,*]

[a] Deparment of Medical Imaging Center, The Affiliated Huaian NO.1 People's Hospital of Nanjing Medical University, Huaian, Jiangsu, China
[b] Deparment of Medical Imaging, Huaian Hospital Affiliated to Xuzhou Medical University, Huaian, Jiangsu, China

ARTICLE INFO

ABSTRACT

Objective: To investigate the effectiveness of a multimodal deep learning model in predicting tumor budding (TB) grading in rectal cancer (RC) patients.

Materials and methods: A retrospective analysis was conducted on 355 patients with rectal adenocarcinoma from two different hospitals. Among them, 289 patients from our institution were randomly divided into an internal training cohort (n = 202) and an internal validation cohort (n = 87) in a 7:3 ratio, while an additional 66 patients from another hospital constituted an external validation cohort. Various deep learning models were constructed and compared for their performance using T1CE and CT-enhanced images, and the optimal models were selected for the creation of a multimodal fusion model. Based on single and multiple factor logistic regression, clinical N staging and fecal occult blood were identified as independent risk factors and used to construct the clinical model. A decision-level fusion was employed to integrate these two models to create an ensemble model. The predictive performance of each model was evaluated using the area under the curve (AUC), DeLong's test, calibration curve, and decision curve analysis (DCA). Model visualization Gradient-weighted Class Activation Mapping (Grad-CAM) was performed for model interpretation.

Results: The multimodal fusion model demonstrated superior performance compared to single-modal models, with AUC values of 0.869 (95% CI: 0.761–0.976) for the internal validation cohort and 0.848 (95% CI: 0.721–0.975) for the external validation cohort. N-stage and fecal occult blood were identified as clinically independent risk factors through single and multivariable logistic regression analysis. The final ensemble model exhibited the best performance, with AUC values of 0.898 (95% CI: 0.820–0.975) for the internal validation cohort and 0.868 (95% CI: 0.768–0.968) for the external validation cohort.

Conclusion: Multimodal deep learning models can effectively and non-invasively provide individualized predictions for TB grading in RC patients, offering valuable guidance for treatment selection and prognosis assessment.

---

* Corresponding author. The Affiliated Huaian NO.1 People's Hospital of Nanjing Medical University, No. 1 West Huanghe Road, Huaian, 223300, Jiangsu, China.
 E-mail address: hybgj0451@163.com (G. Bai).
[1] Ziyan Liu and Jianye Jia contributed equally to this work.

1. Introduction

Colorectal cancer (CRC) is one of the more common gastrointestinal tumors, with rectal cancer (RC) accounting for about one-third of these cases. In 2020, CRC ranked second in global cancer mortality, and the total incidence of CRC in China also ranked second among all cancers, indicating a significant disease burden both globally and in China [1]. With the advancement of endoscopic technology, early-stage RC has become a suitable candidate for endoscopic treatment. However, whether additional radical surgery is necessary after endoscopic treatment for these patients remains a clinical challenge [2]. Currently, the presence of tumor budding (TB) has become one of the evaluation criteria for additional radical surgery after endoscopic treatment in early-stage RC patients [3]. Pathological studies have shown that tumor budding, defined as the presence of single or clusters of fewer than five invasive tumor cells at the forefront of the tumor stroma, indicates the potential for tumor invasion and metastasis, serving as a harbinger process for distant metastasis [4], which is instructive in suggesting early metastasis of RC. Tumor budding is an emerging prognostic biomarker in CRC and other solid tumors. There is ample evidence supporting the prognostic value of tumor budding. TB has been listed as an important adverse prognostic factor for CRC patients in both the TNM staging system (2017) and the WHO classification scheme (2019) [5,6], underscoring its diagnostic importance in assessing RC tumor treatment response and prognosis [7,8]. Studies have indicated that RC patients with TB have a higher incidence of local recurrence and liver metastases [9], and a high-grade TB is significantly associated with shorter overall survival (OS) and disease-free survival (DFS) ($P < 0.05$) [10]. Therefore, clear grading of TB can assist clinicians in determining whether early adjunctive chemotherapy is needed postoperatively in RC patients to improve prognosis [11,12].

Currently, the assessment of tumor budding (TB) in pathological examinations employs hematoxylin and eosin (HE) staining or immunohistochemistry (IHC) methods. These approaches have limitations, including being invasive and failing to comprehensively represent the full scope of the lesion. Additionally, TB is not a mandatory item in pathological tests, which further limits its clinical utility. Therefore, a non-invasive and holistic tumor sampling technique is crucial for timely, comprehensive, and accurate prediction of TB grading in rectal cancer (RC) patients. Recent studies indicate that multimodal integration, which leverages unique information from different imaging modalities, offers a more comprehensive view of tumor biological behavior [13]. Moreover, deep transfer learning has been widely applied in colorectal cancer tumors. Pai and Liu [14,15] used deep learning algorithms to analyze pathological images of colorectal cancer, demonstrating the potential of deep learning in identifying and quantifying the histopathological features of colorectal cancer tissues. Hence, our study innovatively constructs a multimodal deep transfer learning model based on CT and MRI images, aimed at investigating the value of this model in predicting the TB grading in rectal cancer (RC) patients.

2. Materials and methods

2.1. Patients

In a retrospective study, a total of 289 patients who underwent rectal CT and MRI enhancement examinations at Institution 1 from November 2019 to December 2022 were identified. These patients were randomly divided into an internal training cohort (n = 202) and an internal validation cohort (n = 87) using a 7:3 ratio. Inclusion criteria were: (1) pathologically confirmed rectal adenocarcinoma with an assessment of TB grading; (2) Underwent rectal MRI and CT enhanced scans within one week before surgery, with a maximum interval of three days between the two scans; (3) availability of complete clinical and pathological data. Exclusion criteria included: (1) any preoperative treatment (such as radiotherapy, chemotherapy, or immunotherapy); (2) concurrent other malignant
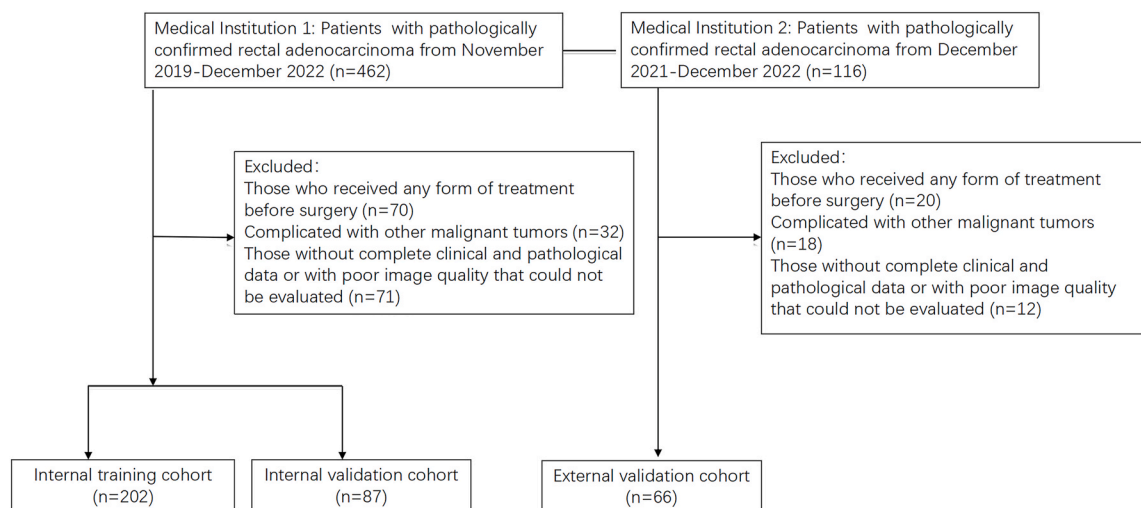

**Fig. 1.** Flow chart of the patients' recruitment pathway.

tumors; (3) missing MRI sequences or poor image quality. Ultimately, 208 patients with rectal adenocarcinoma were included in the study. An external validation cohort comprised 66 eligible patients with rectal adenocarcinoma from Institution 2 (Fig. 1). The study protocol was approved by the Ethics Committee of the Affiliated Huaian No. 1 People's Hospital of Nanjing Medical University (KY-2022-045-01). Patient consent was waived by the institutional review board due to the retrospective and anonymized nature of the data analysis. The study was conducted in accordance with the guidelines of the medical imaging artificial intelligence examination checklist [16].

2.2. Pathological assessment

In accordance with the tumor budding (TB) interpretation and counting criteria established by the International Tumor Budding Consensus Conference (ITBCC), tumor budding was identified as cell clusters containing tumor cells scattered at the tumor front [17]. A microscope with an eyepiece field diameter of 20 mm and an observation area of 0.785 mm^2 was used. Tumor tissue samples from patients were collected, and the hotspot areas where TB was most prevalent were first identified under low magnification. Subsequently, the maximum number of budding instances in the TB hotspot area was counted under a 200x field of view, with at least two pathological slides observed per case. According to the ITBCC grading criteria, the categories were classified as: 0–4 buds (low-grade budding, Bd1), 5–9 buds (intermediate-grade budding, Bd2), and ≥10 buds (high-grade budding, Bd3). In this study, the TB categories were divided into low-intermediate (Bd1+2) and high-grade (Bd3) based on the aforementioned standards. The grading of TB was jointly assessed by two senior pathologists using a multi-headed microscope throughout the study.

2.3. Image acquisition and Region of Interest segmentation

Images were acquired using a 1.5 T magnetic resonance imaging (MRI) scanner (Siemens, Avanto or Aera, Germany), a 3.0 T MRI scanner (Siemens, Verio or Spectra, Germany), and a Siemens SOMATOM Definition dual-source CT scanner from Germany. The specific parameters for the machine scans are detailed in Supplementary Material 1. CT Enhanced Scan: A high-pressure injector is used to inject non-ionic contrast agent iohexol through the elbow vein at a dose of 1.5 ml/kg, at a rate of 3 ml/s, with scanning starting 30 s after contrast agent injection. MRI Enhanced Scan: Contrast agent Gd-DTPA is injected at a dose of 0.1 mmol/kg and a rate of 2.0–3.0 ml/s, followed by an equal volume of saline at the same flow rate, with a dynamic scan performed for 5 min. The T1CE and CT-enhanced images were both standardized and resampled to 1 mm × 1 mm × 1 mm voxels, and subsequently normalized to a range of 0–1 using min-max normalization for each patient's image intensity. Axial images of both T1CE and CT enhancements were imported into the open-source software ITK-SNAP (v.3.8.0, http://www.itksnap.org), where its Registration functionality was utilized for the initial automatic registration of CT and MRI images. During the automatic registration process, manual adjustments were made if errors were detected to ensure precise alignment of the images. The regions of interest (ROI) in the images were manually segmented by two radiologists with 5 and 8 years of experience in abdominal-pelvic MRI diagnosis, who were blinded to the patients' conditions.
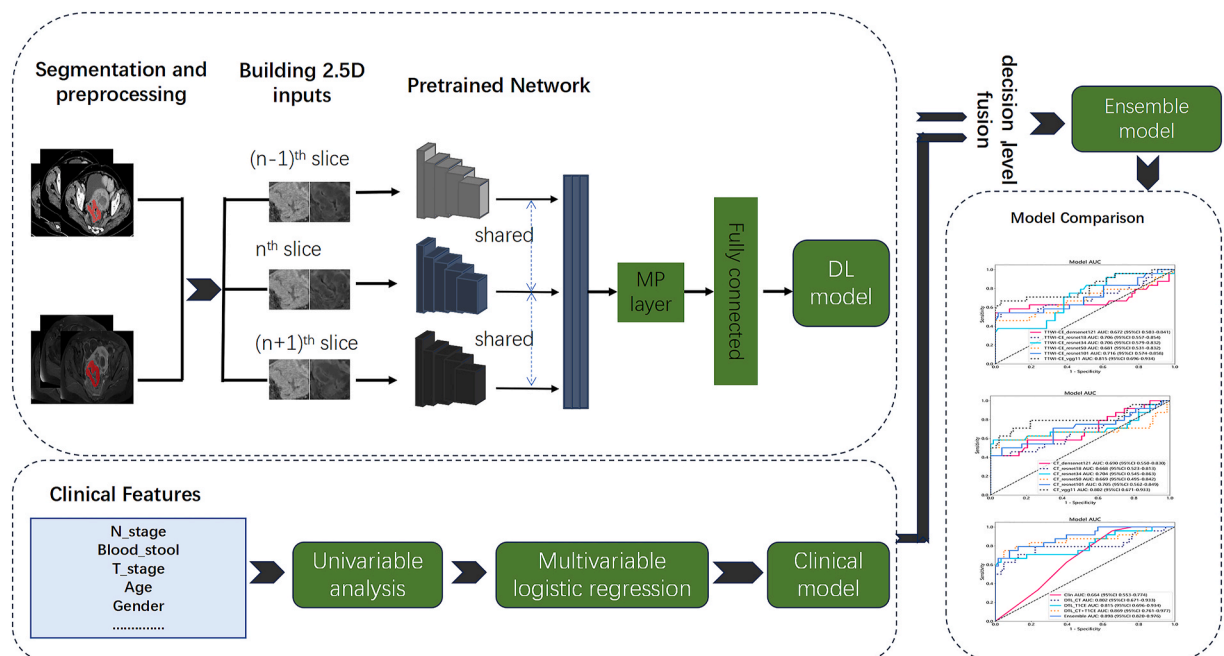
Fig. 2. The schematic workflow of model development.A DLmodel based on 2.5D inputs, and a clinical model were constructed respectively,and the two models were fused through a decision-level fusion (An Ensemble model). DL = deep learning; 2.5D = 2.5 dimension.

The lesion's entire volume was segmented manually on axial images, with adjustments made on coronal and sagittal images to achieve a higher accuracy of ROI. Any discrepancies were resolved through discussion to reach a consensus.

2.4. Deep learning model analysis

Based on the 3D segmentation mask of the tumor, the Region of Interest (ROI) was selected from the original image, focusing on the largest cross-section and its adjacent layers, to create a three-channel 2.5D input method. This approach balances the simplicity of 2D imaging with the spatial information of 3D images. The transfer learning models employed in this study include DenseNet 121, ResNet 18, ResNet 34, ResNet 50, ResNet 101, and Vgg11. Each model was pre-trained on the ImageNet dataset to obtain initial weight values, aiding in rapid convergence and enhancing the models' generalization capabilities on medical images. Prior to training, all such 2.5D inputs were resized to 224×224 pixels. The training process comprised forward computation and backpropagation; additionally, dynamic data augmentation techniques such as random horizontal flipping and cropping were applied to effectively enhance the models' robustness and generalization ability. A focal loss function was utilized to address class imbalance issues. The Stochastic Gradient Descent (SGD) optimizer was used for updating model parameters, starting with an initial learning rate of 0.01. This rate was decreased using a cosine annealing algorithm over 100 epochs and 1800 iteration steps, with a batch size of 128, ensuring ample iterations for the models to learn and adjust weights. Subsequently, we constructed and compared the efficacy of different deep learning models based on CT and MRI modalities, and the optimal models were combined into a multimodal fusion model using decision-level fusion (Fig. 2). The best weights for each model were determined using the predicted probabilities of the individual

Table 1
Clinical baseline information for patients.

Characteristic	Training Cohort			Internal Validation Cohort			External Validation Cohort		
	Bd (1 + 2) (n = 142)	Bd (3) (n = 60)	p-value	Bd (1 + 2) (n = 63)	Bd (3) (n = 24)	p-value	Bd (1 + 2) (n = 48)	Bd (3) (n = 18)	p-value
Sex			0.500			0.637			0.641
Male	81 (57.0%)	38 (63.3%)		47 (74.6%)	16 (66.7%)		29 (60.40%)	6 (33.30%)	
Female	61 (43.0%)	22 (36.7%)		16 (25.4%)	8 (33.3%)		19 (39.60%)	12 (66.70%)	
Age (years)	65.35 ± 10.38	66.07 ± 10.60	0.752	64.46 ± 9.11	65.38 ± 9.17	0.771	64.60 ± 11.30	65.50 ± 11.70	0.791
Size (cm)	4.00 (3.00, 5.00)	4.00 (3.25, 5.00)	0.772	4.00 (2.50, 4.50)	4.00 (3.50,4.75)	0.208	3.75 (2.75,5.00)	3.35 (2.50,4.00)	0.381
Distance (cm)	8.00 (5.00, 10.00)	8.0 (5.00, 10.00)	0.757	8.00 (5.00, 10.00)	6.50 (5.00, 10.00)	0.881	8.00 (5.00,10.00)	6.00 (5.00,8.00)	0.082
T stage			0.314			0.974			0.555
T1	8 (5.6%)	0 (0.0%)		0 (0.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	
T2	44 (31.0%)	19 (31.7%)		21 (33.3%)	8 (33.3%)		20 (41.7%)	8 (44.4%)	
T3	88 (62.0%)	40 (66.7%)		40 (63.5%)	15 (62.5%)		25 (52.1%)	10 (55.6%)	
T4	2 (1.4%)	1 (1.7%)		2 (3.2%)	1 (4.2%)		3 (6.3%)	0 (0.0%)	
N stage			<0.001			0.680			0.693
N0	74 (52.1%)	13 (21.7%)		27 (42.9%)	8 (33.3%)		12 (25.0%)	3 (16.7%)	
N1	29 (20.4%)	12 (20.0%)		14 (22.2%)	7 (29.2%)		12 (25.0%)	6 (33.3%)	
N2	39 (27.5%)	35 (58.3%)		22 (34.9%)	9 (37.5%)		24 (50.0%)	9 (50.0%)	
Stomachache			0.819			0.627			0.313
Yes	48 (33.8%)	22 (36.7%)		26 (41.9%)	8 (33.3%)		20 (41.7%)	10 (42.9%)	
No	94 (66.2%)	38 (63.3%)		36 (58.1%)	16 (66.7%)		28 (58.3%)	12 (57.1%)	
Ventosity			0.851			0.791			0.362
Yes	13 (9.2%)	5 (8.3%)		11 (17.7%)	3 (12.5%)		10 (20.8%)	2 (11.1%)	
No	129 (90.8%)	55 (91.7%)		51 (82.3%)	21 (87.5%)		38 (79.2%)	16 (88.9%)	
Tenesmus			0.902			0.720			0.952
Yes	32 (22.5%)	14 (11.9%)		14 (22.6%)	7 (29.2%)		11 (22.9%)	4 (22.2%)	
No	110 (77.5%)	46 (76.7%)		48 (77.4%)	17 (70.8%)		37 (77.1%)	14 (77.8%)	
Bloody_stool			0.001			0.003			0.327
Yes	89 (62.7%)	52 (86.7%)		40 (64.5%)	23 (95.8%)		42 (87.5%)	14 (77.8%)	
No	53 (37.3%)	8 (13.3%)		22 (35.5%)	1 (4.2%)		6 (12.5%)	4 (22.2%)	
Time			0.860			0.924			0.424
Yes	58 (40.8%)	23 (38.3%)		20 (32.3%)	8 (33.3%)		9 (18.8%)	5 (27.8%)	
No	84 (59.2%)	37 (61.7%)		42 (67.7%)	16 (66.7%)		39 (81.3%)	13 (72.2%)	
Character			0.571			0.423			0.421
Yes	89 (62.7%)	39 (65.0%)		31 (50.0%)	15 (62.5%)		24 (50.0%)	7 (38.9%)	
No	53 (37.3%)	21 (35.0%)		31 (50.0%)	9 (37.5%)		24 (50.0%)	11 (61.1%)	
CA19-9 (U/mL)	9.92 (5.71,14.90)	11.35 (7.08,23.65)	0.067	8.28 (5.17,20.87)	10.29 (6.65,22.25)	0.342	10.25 (5.81,16.45)	7.78 (5.07,17.70)	0.579
CA50 (U/mL)	6.44 (3.98,9.65)	7.50 (4.93, 12.86)	0.051	5.98 (3.90, 11.65)	6.12 (4.58,15.83)	0.395	6.33 (3.97,9.27)	5.72 (3.89,10.30)	0.908
CEA (ng/mL)	2.91 (1.83,6.53)	3.94 (2.07,10.80)	0.096	3.49 (2.27,7.86)	3.48 (1.99,4.46)	0.344	3.52 (2.04,9.90)	2.85 (1.96,4.10)	0.190

Note: Data are mean ± standard deviation or median with interquartile range and numbers in parenthese. CA19-9: Carbohydrate antigen19-9; CA50: Carbohydrate antigen50; CEA: Carcinoembryonic antigen.

models on the training data as inputs for logistic regression. Once the optimal weights were established, a soft voting scheme was employed to integrate the models (Equation in Supplementary Materials 2).

2.5. Statistical analysis

Statistical analyses were conducted using R software (version 4.1.3: www.R-project.org) and SPSS software (version 25.0, IBM, Armonk, NY, USA). Depending on the distribution of variables, quantitative data with a normal distribution were expressed as mean $\pm$ standard deviation and analyzed using independent sample t-tests. Quantitative data with a non-normal distribution were expressed as median (interquartile range) and analyzed using the Mann-Whitney U test. Categorical data were analyzed using the Chi-square test or Fisher's exact test and expressed in percentages.We utilized univariate logistic regression analysis to assess the initial associations between each clinical indicator and the study outcomes. Subsequently, to determine independent risk factors, we further employed multivariable logistic regression analysis. The analysis procedure selected forward selection for variable screening. A p-value of <0.05 was considered statistically significant. The predictive performance of each model was evaluated using the area under the curve (AUC), DeLong's test, calibration curves, and decision curve analysis (DCA). The F1 score was calculated to assess and compare the predictive performance of each model (Supplementary Material 3). Gradient-weighted Class Activation Mapping (Grad-CAM) was used for model interpretation.

3. Results

3.1. Clinical data and model establishment

A total of 355 patients were included in the study, with 253 patients (71.3%) in the Bd (1 + 2) group and 102 patients (28.7%) in the Bd (3) group. The clinical characteristics are as follows (Table 1): there was no statistical significance in patients' gender, age, tumor diameter, distance from the anal margin, clinical T stage, symptoms of abdominal pain, bloating, tenesmus, frequency and nature of bowel movements, and tumor markers (CA199, CA50, CEA). Through univariate and multivariate logistic regression analyses, two independent predictive factors (clinical N stage and bloody stool) were identified (Table 2). A clinical model was constructed, and the AUC values for the internal and external validation cohorts were 0.664 (95% CI: 0.554–0.773) and 0.634 (95% CI: 0.504–0.763), respectively.

3.2. Fusion and ensemble model establishment and evaluation

In the deep learning models constructed based on CT images, the Vgg11 model demonstrated the best performance, with AUC values for the internal and external validation cohorts being 0.802 (95% CI: 0.671–0.932) and 0.758 (95% CI: 0.609–0.907), respectively. Similarly, in the deep learning models based on MRI images, the Vgg11 model also showed outstanding performance, with AUC values for the internal and external validation cohorts of 0.815 (95% CI: 0.697–0.933) and 0.764 (95% CI: 0.604–0.924), respectively(Supplementary Material 4). Based on these findings, we performed a decision-level fusion of the Vgg11 models from CT and MRI to construct a multimodal fusion model. Excitingly, the performance of the fusion model significantly surpassed that of the single modality imaging approach, with AUC values for the internal and external validation cohorts of 0.869 (95% CI: 0.761–0.976) and 0.848 (95% CI: 0.721–0.975), respectively. Further, we combined the fusion model with clinical independent risk factors to construct an ensemble model. The results indicated that the AUC values for the internal and external validation cohorts were 0.898 (95% CI: 0.820–0.975) and 0.868 (95% CI: 0.768–0.968), respectively (Table 3). Compared to the individual clinical and deep learning models, the ensemble model showed significant improvements in accuracy, sensitivity, specificity, and predictive performance. The calibration curve was evaluated using the Hosmer-Lemeshow (HL) test, which showed good consistency between predicted and actual outcomes. Decision curve analysis (DCA) indicated that within the risk threshold, the ensemble model provided higher overall net benefit and clinical decision-making efficacy compared to the clinical model and deep learning models (Fig. 3).

3.3. Model interpretation

Gradient-weighted Class Activation Mapping (Grad-CAM) provides an intuitive visualization of the model's specific decisions through heatmaps, where different colors represent the convolutional neural network's focus on different regions of the input image [18]. By overlaying the heatmap of the last convolutional layer onto the original image, Grad-CAM helps identify the image regions on which the model relies when diagnosing diseases, thereby offering crucial insights into its decision-making process. This is particularly

Table 2
Independent risk factors were screened by univariable and multivariable logistic regression analysis.

Factors	Univariable logistic regression		Multivariable logistic regression	
	OR (95% CI)	p value	OR (95% CI)	p value
N_stage	2.255 (1.559–3.261)	< 0.001	2.173 (1.490–3.171)	< 0.001
Blood_stool	3.871 (1.708–8.774)	0.001	3.51 (1.507–8.173)	0.004

Table 3
F1-score, sensitivity, specificity, and AUC values in the five models.

| model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	F1-score	Task
Clinic	0.673	0.731	0.662–0.799	0.667	0.676	0.548	Training cohort
Clinic	0.609	0.664	0.554–0.773	0.625	0.603	0.469	Internal Validation cohort
Clinic	0.576	0.634	0.504–0.763	0.500	0.604	0.391	External Validation cohort
DTL_CT	0.822	0.821	0.755–0.886	0.783	0.838	0.723	Training cohort
DTL_CT	0.816	0.802	0.671–0.933	0.667	0.873	0.667	Internal Validation cohort
DTL_CT	0.833	0.758	0.609–0.907	0.389	1.000	0.560	External Validation cohort
DTL_T1CE	0.906	0.837	0.761–0.913	0.683	1.000	0.812	Training cohort
DTL_T1CE	0.874	0.815	0.696–0.933	0.625	0.968	0.732	Internal Validation cohort
DTL_T1CE	0.864	0.764	0.603–0.924	0.500	1.000	0.667	External Validation cohort
DTL_CT + T1CE	0.871	0.936	0.897–0.974	0.867	0.873	0.800	Training cohort
DTL_CT + T1CE	0.885	0.869	0.761–0.977	0.708	0.952	0.773	Internal Validation cohort
DTL_CT + T1CE	0.848	0.848	0.721–0.976	0.722	0.896	0.722	External Validation cohort
Ensemble	0.886	0.954	0.918–0.991	0.917	0.873	0.827	Training cohort
Ensemble	0.862	0.898	0.820–0.975	0.708	0.952	0.773	Internal Validation cohort
Ensemble	0.848	0.868	0.768–0.968	0.722	0.896	0.722	External Validation cohort

significant in medical image analysis, where researchers can understand how the model identifies disease features by observing these heatmaps, providing valuable insights into its decision-making process. Moreover, Grad-CAM not only enhances the transparency of the model but also aids in identifying and correcting model biases. For instance, if the model erroneously focuses on irrelevant areas, it may indicate issues with the training data or require adjustments in the learning approach. Hence, Grad-CAM offers an intuitive way to understand and explain complex neural network decisions, especially in the critical field of medical image processing. Through this technique, we can enhance the credibility and practicality of the model. Fig. 4 illustrates the comparative results of the original images and overlays of heatmaps from different deep learning models. The inward-concentrated red regions indicate activation, suggesting that the model is particularly attentive to these areas.

4. Discussion

Recent years have seen the prediction analysis of combinations of multiple modal variable features as a more effective method to assist clinical practice, providing more comprehensive information support for clinical diagnosis and treatment [19]. In our study, we used decision-level fusion to construct anensemble model combining the clinical model and the fusion model. This model demonstrated superior diagnostic ability over single models, exhibiting outstanding diagnostic performance in both internal and external validation cohorts, indicating its strong applicability. Moreover, using data from two centers for validation also enhanced the stability and universality of our study results.

Tumor budding, as an emerging prognostic biomarker in colorectal cancer, adheres to the principle that 'higher budding levels correlate with worse clinical outcomes,' regardless of clinical situation or tumor type [3]. Therefore, clear grading of TB is helpful for clinicians to administer precise treatment for patients. Currently, traditional radiomics methods have been used to construct predictive models for rectal cancer tumor budding. Li [20] developed an imaging-based model using multi-sequence MRI to predict tumor budding grades in rectal cancer patients, with AUC values of 0.875 (95% CI: 0.752–0.951) for internal validation and 0.796 (95% CI: 0.702–0.871) for external validation. Furthermore, Peng [21] added a clinical model to the study to construct a combined model, with an AUC value of 0.891 (95% CI: 0.800–0.981) for validation. However, the predictive performance of theensemble model constructed in our study surpasses these levels, with AUC values of 0.898 (95% CI: 0.820–0.975) and 0.868 (95% CI: 0.768–0.968) for internal and external validation cohorts, respectively. This improvement may be attributed to the use of deep transfer learning models, which, unlike handcrafted features, can automatically learn complex features directly from the raw pixels of input images for end-to-end classification and prediction [22]. Additionally, the inclusion of multimodal data, allowing for information extraction from different types, may also contribute, enabling a more comprehensive and holistic understanding of the biological information of lesions [23]. Pai and Liu [14,15]have both utilized deep learning methods to analyze pathological images of rectal cancer, demonstrating the potential of deep learning in identifying and quantifying histopathological features including tumor budding. Although these studies emphasize the application value of deep learning in pathological image analysis, our study extends the application scope of deep learning techniques to the analysis of multimodal data based on MRI and CT images, yielding promising results. In comparison to their research, we not only focus on pathological images but also analyze MRI and CT images using deep learning techniques, allowing us to make predictions without the need for actual pathological samples, thereby providing the possibility for non-invasive predictions.

In our study, we selected only the T1CE sequence from MRI images, based on Li and Peng's multi-sequence MRI research, which found that enhanced sequence models showed better predictive performance, in line with our clinical practice. Delong tests(Supplementary Material 5) revealed that the T1CE model performed slightly better than CT in both internal and external validation cohorts, but the difference was not statistically significant. This minor difference could stem from MRI's advantage in soft tissue resolution. Although the difference between the fusion model and single modality models was not statistically significant, considering improvements in sensitivity and performance, the advantages of multimodal models remain significant, especially in analyzing complex tumor features, such as differentiating subtypes or assessing tumor invasion into surrounding tissues [24,25]. Therefore, while single modality imaging methods can be sufficiently predictive in some cases, combining the advantages of CT and MRI provides
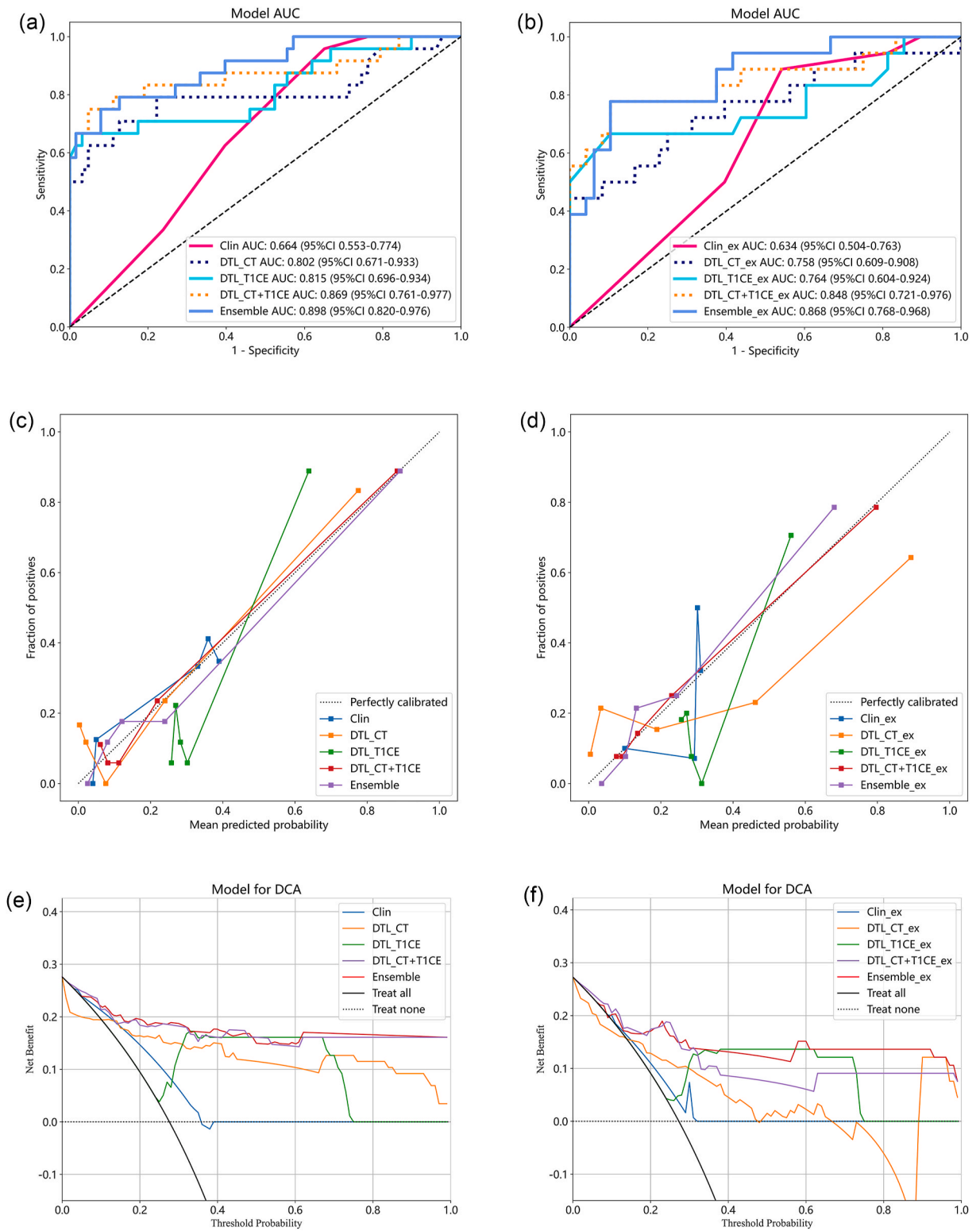
Fig. 3. A, B: Comparison of the AUC of each model in the internal and external validation cohorts; C, D: Calibration curves of each model in the internal and external validation cohorts; E, F: DCA curves of each model in the internal and external validation cohorts.
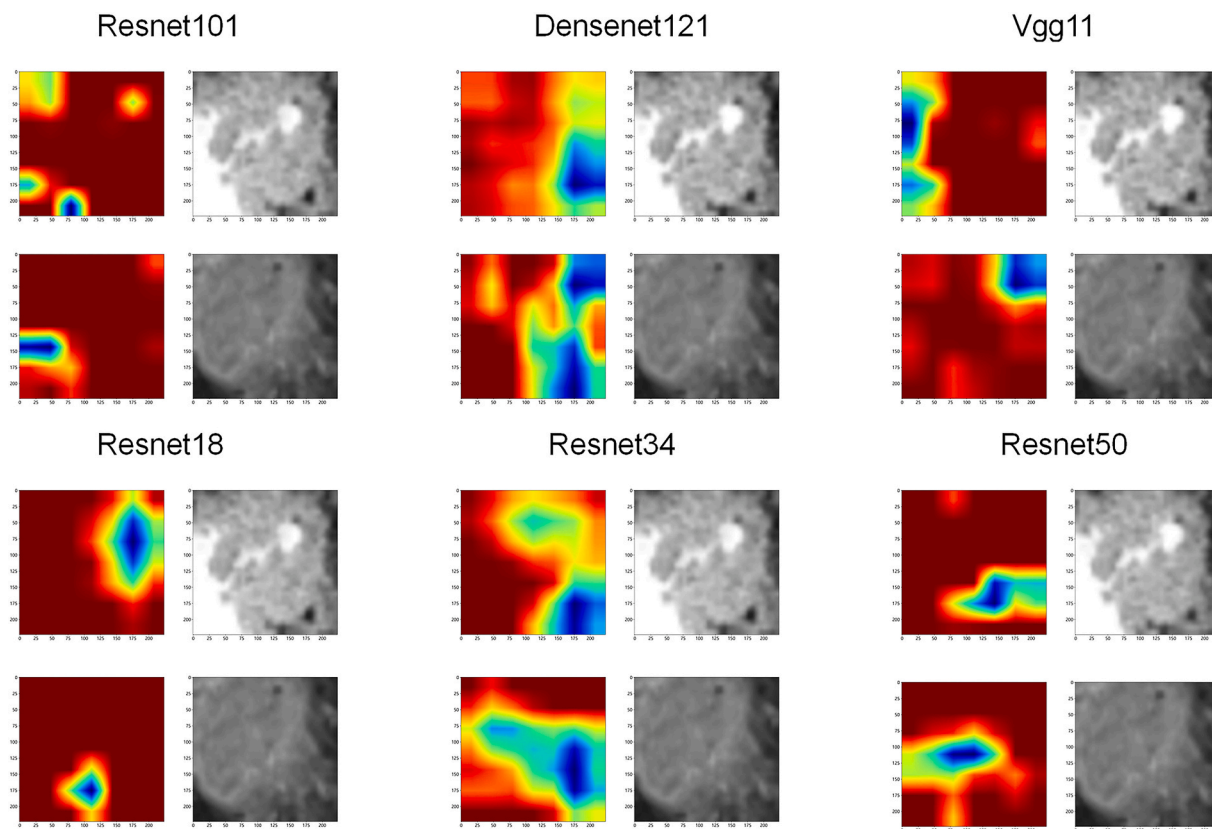
Fig. 4. Grad-CAM Heatmaps, Highlighting Key Areas and Original Lesion Regions Predicted by the Model. The areas focused on by different deep learning models in the analysis of CT and MRI images. Regions of the lesion predominantly in red indicate the highest activation level by the model in these parts. Areas predominantly in blue represent lower activation levels, suggesting that these regions have a lesser impact on the model's diagnostic decision-making. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

richer data, offering more accurate predictions in complex clinical situations [26–28]. Moreover, we must recognize that multi-sequence parameters in MRI images might provide a more comprehensive view of the tumor [29,30], aiding in more accurately identifying specific tumor characteristics, which has not been fully explored in our study. Future research should explore the combined use of different imaging technologies. This finding also suggests that clinicians, when choosing imaging techniques, should consider not only availability and cost-effectiveness but also the comprehensive advantages of different imaging methods to achieve optimal clinical decision-making.

This study employed six transfer learning models (including DenseNet 121, ResNet 18, ResNet 34, ResNet 50, ResNet 101, and Vgg11) to predict the grading of tumor budding in patients with rectal cancer. Among these models, the Vgg11 model exhibited stronger generalization and predictive capabilities compared to other classic CNNs. The performance differences between various CNN models could be attributed to differences in their internal network architectures. The Vgg11 model, with its relatively simple structure and fewer parameters, helps reduce the risk of overfitting in small datasets or simple tasks [31,32], which could be advantageous in practical clinical applications. Furthermore, the Grad-CAM visualization of the Vgg11 model effectively highlights specific features within rectal cancer images. These features predominantly manifest as morphological heterogeneity in the tumor area, such as the indistinctness of tumor margins and irregularity in shape. Additionally, the heterogeneity of the internal structure of the tumor is also recognized by the model as a significant feature. This confirms the model's effectiveness in identifying and analyzing key tumor characteristics. In this manner, the model not only demonstrates its diagnostic capabilities on a technical level but also provides physicians with a more intuitive means of understanding the pathological characteristics of the tumor.

This study has certain limitations: (1) It is a retrospective study and may contain inherent selection biases. Further research involving more centers and prospective experiments is necessary. (2) To avoid overfitting due to too many training parameters, we only used 2.5D inputs for the deep learning models, inevitably missing some image information. Future studies might need to increase the sample size and implement 3D inputs. (3) This study only considered the T1CE sequence for MRI images. Future studies incorporating more sequence parameters for MRI may achieve better predictive results.

In conclusion, the multimodal deep learning model constructed in this study can effectively and non-invasively perform individualized prediction of TB grading in RC patients. This has significant implications for guiding the choice of treatment and prognosis

assessment for patients.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because the subjects did not provide written consent for their data to be publicly shared.

CRediT authorship contribution statement

Ziyan Liu: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Jianye Jia:** Writing – review & editing, Methodology, Data curation. **Fan Bai:** Methodology, Formal analysis, Data curation. **Yuxin Ding:** Methodology, Data curation. **Lei Han:** Data curation. **Genji Bai:** Writing – review & editing, Project administration, Conceptualization.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28769.

References

[1] R.L. Siegel, K.D. Miller, A. Goding Sauer, et al., Colorectal cancer statistics, 2020, CA Cancer J Clin 70 (2020) 145–164.
[2] N. Dekkers, H. Dang, J. van der Kraan, et al., Risk of recurrence after local resection of T1 rectal cancer: a meta-analysis with meta-regression, Surg. Endosc. 36 (2022) 9156–9168.
[3] A. Lugli, I. Zlobec, M.D. Berger, et al., Tumour budding in solid cancers, Nat. Rev. Clin. Oncol. 18 (2021) 101–115.
[4] A.B. Benson, A.P. Venook, M.M. Al-Hawary, et al., Rectal cancer, version 2.2022, NCCN clinical practice guidelines in oncology, J. Natl. Compr. Cancer Netw. 20 (2022) 1139–1167.
[5] J. Doescher, J.A. Veit, T.K. Hoffmann, HNO 65 (2017) 956–961 [The 8th edition of the AJCC Cancer Staging Manual : Updates in otorhinolaryngology, head and neck surgery].
[6] I.D. Nagtegaal, R.D. Odze, D. Klimstra, et al., The 2019 WHO classification of tumours of the digestive system, Histopathology 76 (2020) 182–188.
[7] G. Karayannopoulou, E. Panteris, J. Kanitakis, Tumour budding is an independent predictive factor of cutaneous squamous-cell carcinoma aggressiveness, Anticancer Res. 40 (2020) 2695–2699.
[8] I. Trotsyuk, H. Sparschuh, A.J. Müller, et al., Tumor budding outperforms ypT and ypN classification in predicting outcome of rectal cancer after neoadjuvant chemoradiotherapy, BMC Cancer 19 (2019) 1033.
[9] T. Ito, N. Kuriyama, Y. Kozuka, et al., High tumor budding is a strong predictor of poor prognosis in the resected perihilar cholangiocarcinoma patients regardless of neoadjuvant therapy, showing survival similar to those without resection, BMC Cancer 20 (2020) 209.
[10] H. Dawson, F. Galuppini, P. Träger, et al., Validation of the International Tumor Budding Consensus Conference 2016 recommendations on tumor budding in stage I-IV colorectal cancer, Hum. Pathol. 85 (2019) 145–151.
[11] J.K. Shin, Y.A. Park, J.W. Huh, et al., Tumor budding as a prognostic marker in rectal cancer patients on propensity score analysis, Ann. Surg Oncol. 28 (2021) 8813–8822.
[12] F. Chen, S. Zhang, X. Ma, et al., Prediction of tumor budding in patients with rectal adenocarcinoma using b-value threshold map, Eur. Radiol. 33 (2023) 1353–1363.
[13] M.A. Azam, K.B. Khan, S. Salahuddin, et al., A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics, Comput. Biol. Med. 144 (2022) 105253.
[14] R.K. Pai, D. Hartman, D.F. Schaeffer, et al., Development and initial validation of a deep learning algorithm to quantify histological features in colorectal carcinoma including tumour budding/poorly differentiated clusters, Histopathology 79 (2021) 391–405.
[15] S. Liu, Y. Zhang, Y. Ju, et al., Establishment and clinical application of an artificial intelligence diagnostic platform for identifying rectal cancer tumor budding, Front. Oncol. 11 (2021) 626626.
[16] J. Mongan, L. Moy, C.E. Kahn Jr., Checklist for artificial intelligence in medical imaging (CLAIM): a Guide for authors and reviewers, Radiol Artif Intell 2 (2020) e200029.
[17] A. Lugli, R. Kirsch, Y. Ajioka, et al., Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016, Mod. Pathol. 30 (2017) 1299–1311.
[18] H. Zhang, K. Ogasawara, Grad-CAM-based explainable artificial intelligence related to medical text processing, Bioengineering (Basel) 10 (2023) 1070.
[19] Y. Cui, W. Yang, J. Ren, et al., Prognostic value of multiparametric MRI-based radiomics model: potential role for chemotherapeutic benefits in locally advanced rectal cancer, Radiother. Oncol. 154 (2021) 161–169.

[20] Z. Li, F. Chen, S. Zhang, et al., The feasibility of MRI-based radiomics model in presurgical evaluation of tumor budding in locally advanced rectal cancer, Abdom Radiol (NY) 47 (2022) 56–65.

[21] L. Peng, D. Wang, Z. Zhuang, et al., Preoperative noninvasive evaluation of tumor budding in rectal cancer using multiparameter MRI radiomics, Acad. Radiol. S1076–6332 (23) (2023), 00656-00656 [pii].

[22] K. Wang, X. Lu, H. Zhou, et al., Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study, Gut 68 (2019) 729–741.

[23] F. Mohsen, H. Ali, N. El Hajj, et al., Artificial intelligence-based methods for fusion of electronic health records and imaging data, Sci. Rep. 12 (2022) 17981.

[24] S. Dayarathna, K.T. Islam, S. Uribe, et al., Deep learning based synthesis of MRI, CT and PET: review and analysis, Med. Image Anal. 92 (2023) 103046.

[25] S. Bedrikovetski, N.N. Dudi-Venkata, H.M. Kroon, et al., Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis, BMC Cancer 21 (2021) 1058.

[26] X. Jiang, H. Zhao, O.L. Saldanha, et al., An MRI deep learning model predicts outcome in rectal cancer, Radiology 307 (2023) e222223.

[27] L. Wan, J. Hu, S. Chen, et al., Prediction of lymph node metastasis in stage T1-2 rectal cancers with MRI-based deep learning, Eur. Radiol. 33 (2023) 3638–3646.

[28] W. Cao, H. Hu, J. Guo, et al., CT-based deep learning model for the prediction of DNA mismatch repair deficient colorectal cancer: a diagnostic study, J. Transl. Med. 21 (2023) 214.

[29] G. Zhang, L. Chen, A. Liu, et al., Comparable performance of deep learning-based to manual-based tumor segmentation in KRAS/NRAS/BRAF mutation prediction with MR-based radiomics in rectal cancer, Front. Oncol. 11 (2021) 696706.

[30] Z. Shu, D. Mao, Q. Song, et al., Multiparameter MRI-based radiomics for preoperative prediction of extramural venous invasion in rectal cancer, Eur. Radiol. 32 (2022) 1002–1013.

[31] N. Fujima, V.C. Andreu-Arasa, K. Onoue, et al., Utility of deep learning for the diagnosis of otosclerosis on temporal bone CT, Eur. Radiol. 31 (2021) 5206–5211.

[32] M. Zhang, M. Xue, S. Li, et al., Fusion deep learning approach combining diffuse optical tomography and ultrasound for improving breast cancer classification, Biomed. Opt Express 14 (2023) 1636–1646.