



A multi-class classification algorithm based on hematoxylin-eosin staining for neoadjuvant therapy in rectal cancer: a retrospective study

Yihan Wu^{1,2,*}, Xiaohua Liu^{3,*}, Fang Liu⁴, Yi Li^{2,3}, Xiaomin Xiong^{2,3}, Hao Sun⁵, Bo Lin², Yu Li⁴ and Bo Xu^{1,2}

¹ School of Medicine, Chongqing University, Chongqing, China

² Chongqing Key Laboratory of Intelligent Oncology for Breast Cancer, Chongqing University Cancer Hospital, Chongqing, China

³ Bioengineering College of Chongqing University, Chongqing, China

⁴ Department of Pathology, Chongqing University Cancer Hospital, School of Medicine, Chongqing University, Chongqing, China

⁵ Department of Gastrointestinal Cancer Center, Chongqing University Cancer Hospital, Chongqing, China

* These authors contributed equally to this work.

ABSTRACT

Neoadjuvant therapy (NAT) is a major treatment option for locally advanced rectal cancer. With recent advancement of machine/deep learning algorithms, predicting the treatment response of NAT has become possible using radiological and/or pathological images. However, programs reported thus far are limited to binary classifications, and they can only distinguish the pathological complete response (pCR). In the clinical setting, the pathological NAT responses are classified as four classes: (TRG0-3), with 0 as pCR, 1 as moderate response, 2 as minimal response and 3 as poor response. Therefore, the actual clinical need for risk stratification remains unmet. By using ResNet (Residual Neural Network), we developed a multi-class classifier based on Hematoxylin-Eosin (HE) images to divide the response to three groups (TRG0, TRG1/2, and TRG3). Overall, the model achieved the AUC 0.97 at 40× magnification and AUC 0.89 at 10× magnification. For TRG0, the model under 40× magnification achieved a precision of 0.67, a sensitivity of 0.67, and a specificity of 0.95. For TRG1/2, a precision of 0.92, a sensitivity of 0.86, and a specificity of 0.89 were achieved. For TRG3, the model obtained a precision of 0.71, a sensitivity of 0.83, and a specificity of 0.88. To find the relationship between the treatment response and pathological images, we constructed a visual heat map of tiles using Class Activation Mapping (CAM). Notably, we found that tumor nuclei and tumor-infiltrating lymphocytes appeared to be potential features of the algorithm. Taken together, this multi-class classifier represents the first of its kind to predict different NAT responses in rectal cancer.

Submitted 27 December 2022

Accepted 21 April 2023

Published 12 June 2023

Corresponding authors

Yu Li, liyu100@cqu.edu.cn

Bo Xu, xubo731@cqu.edu.cn

Academic editor

Xin Zhang

Additional Information and
Declarations can be found on
page 13

DOI 10.7717/peerj.15408

© Copyright
2023 Wu et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Computational Biology, Gastroenterology and Hepatology, Oncology, Pathology

Keywords Rectal cancer, Neoadjuvant therapy, Deep learning, Multi-class classification algorithms, Pathology

INTRODUCTION

Cancers originated from colon and rectal are the third most commonly diagnosed and the second deadliest worldwide (*Sung et al., 2021*). Rectal cancer is distinctive to colon cancer, owing to its anatomical location, etiological factors and environmental associations. For locally advanced rectal cancer, neoadjuvant therapy (NAT) is a standard preoperative treatment strategy, which is conducive to the preservation of the anal sphincter muscle, to the complete surgical removal of tumor cells, as well as to the reduction of tumor micrometastasis and recurrence (*Benson et al., 2020*). The efficacy of NAT is typically evaluated by the postoperative pathological assessment, reported by the tumor regression grade (TRG) (*Karagkounis et al., 2019; National Health Commission of the People's Republic of China, 2020*). The American Joint Committee on Cancer (AJCC) TRG system (*Kim et al., 2016*) consists of Grades 0, 1, 2, and 3, among which TRG0 is regarded as pathological complete response (pCR), indicating that there are no residual active tumor cells in the primary nidus. TRG1 indicates a moderate response with single or small groups of tumor cells remaining. TRG2 reflects a minimal response with residual cancer outgrown by fibrosis. TRG3 is regarded as a poor response with large areas of residual cancer in the primary lesion, and few or no tumor cells are killed. It is believed that TRG grading has predictive values as there are significant correlations between TRGs and the clinical outcome in the disease (*Belluco et al., 2011; de Campos-Lobato et al., 2011*). Unfortunately, there are less than 30% of patients achieving pCR (*Monson & Arsalanizadeh, 2017*). For patients who cannot achieve it, especially for TRG2 and 3, NAT is not beneficial as it might weaken their immune system and delay other treatment. Therefore, approaches to NAT risk stratification are urgently needed.

Over the years, significant efforts have been invested in identifying risk factors of rectal cancer NAT. On the biological side, several categories of biomarker identifications have been explored. For example, studies have shown that the pretreatment tumor marker CEA level (*Wallin et al., 2013; Das et al., 2007*), HEP2, GELS, and S10A8 protein expression in serum (*Wang et al., 2022*), the number of tumor-infiltrating lymphocytes (TILs) (*Matsutani et al., 2018*), subsets of tumor-infiltrating and peripheral blood immune cells (*Zhu et al., 2022*), inflammatory cancer-associated fibroblasts (iCAFs) (*Nicolas et al., 2022*), and the CD44 mRNA level (*Huh, Lee & Kim, 2014*), can be used as biomarkers to predict treatment response in rectal cancer NAT. Higher TILs, the CD4+/CD8+ T cell ratio, HEP2 and GELS, and lower iCAFs, S10A8 and pretreatment CEA levels are associated with better responses.

In recent years, artificial intelligence (AI) based programs have been explored in identifying features of rectal cancer NAT. Based on traditional machine learning (ML) and deep learning (DL), AI can discover and extract hidden information in images. Therefore it plays an increasingly significant role in various aspects of tumor screening, diagnosis, assessment of efficacy, and prediction of prognosis (*Coudray et al., 2018; Bilal et al., 2021; Zhou et al., 2020; She et al., 2020; Courtiol et al., 2019; Lin et al., 2022*). Among them, studies have used radiological images involving CT (*Kurata et al., 2021*) and MRI (*Liu et al., 2017*) to predict the treatment response to NAT in rectal cancer. In addition to radiological images, pathological images have also been used to extract

features that they might be associated with the therapeutic response. Pathological images contain macroscopic histological features and microscopic cellular information, and these characteristics intuitively provide the essential role of the occurrence and development of the disease. DL can extract many high-dimensional features through various algorithms, which are invisible to human eyes. For example, some algorithms can predict microsatellite status (*Yamashita et al., 2021*) and prognosis (*Kather et al., 2019*) based on pathological images of rectal cancer. In addition, studies (*Zhang et al., 2020*; *Shao et al., 2020*) have utilized ML methods to manually extract features from pathological images of rectal cancer patients to predict the pCR.

Despite these promising results, it is noted that all these studies are aimed at identifying features or biomarkers for the pCR patients. Owing to the multi-category grading system, it is apparent that solely finding pCR patients is not sufficient. As a matter of fact, identifying TRG2 and TRG3 patients, which are less responsive to NAT, might be more important, as NAT do not benefit these patients. Although there is a recent study (*Jepsen et al., 2022*) that has developed a multi-class quantitative evaluation algorithm for pathological images of rectal cancer after neoadjuvant therapy, multi-class prediction algorithms have not been reported.

In the present study, we established a three-class classification method to predict the efficacy of NAT in patients with rectal cancer. We show a robust performance in categorizing TRG01, TRG1/2 and TRG3. In addition, we established a heat map based on the CAM method to analyze the interpretability of pathological images in different treatment responses.

MATERIALS & METHODS

Patient selection

A total of 94 patients who received NAT in Chongqing University Cancer Hospital (CUCH) from 2011 to 2022 were retrospectively enrolled. The criteria for inclusion comprised: (1) Diagnosis of primary rectal cancer through pathological biopsy before NAT; (2) receiving of the NAT regimen based on Capecitabine or 5-FU combined with Oxaliplatin; (3) availability of preoperative pathological biopsy sections; (4) have underwent surgery after NAT and have TRG evaluation results. The criteria for exclusion comprised: (1) Other treatments before NAT; (2) lack of preoperative pathological biopsy section; (3) poor quality of H&E-stained images. This study was approved by the Ethics Committee of Chongqing University Cancer Hospital (Ethics number: CZLS2022185-A).

Pathological assessment of response and data collection

Specimens were prepared onto glass slides and stained with hematoxylin & eosin (H&E). TRG grades were evaluated by two experienced pathologists, and samples with discordant scores were reevaluated until an agreement was reached. In addition, clinical data including the age at diagnosis, gender, tumor/node(T/N) stage, CEA and CA199 levels before treatment, and treatment regime were collected. Due to the institutional policy, the pathological image data are not publicly available.

Image preprocessing

H&E-stained slides were produced at 40× magnification by the KFBio KF-PRO-005-HI digital scanner. There were 133 slides collected. Whole slide images (WSIs) were annotated using QuPath open-source software (version 0.3.2). The annotation of tumor tissues in the WSIs was performed by two pathologists. One pathologist was responsible for delineating, and the other for examining. To reduce the computing consumption, the WSIs and their corresponding annotations were split into overlapping 512 × 512 pixels tiles at 0.238 μm/pixel resolution (equivalent to 40× magnification) and 1 μm/pixel resolution (equivalent to 10× magnification) with overlapping segmentation, which stride was 0.5. Tiles with tissues less than 50% were removed manually.

All tiles were color normalized with the structure-preserving color normalization (SPCN) method ([Vahadane et al., 2016](#)). Different types of data augmentation were applied, including random vertical flip (0.5), random horizontal flip (0.5), and brightness/contrast/saturation/hue changes (1).

Construction and visualization of response classifier

The WSIs from CUCH were split at a ratio of approximately 4:1, with 113 WSIs in the training set and 20 WSIs in the test set. The training set consisted of 166,919 tumor ROI tiles at 40× magnification. There were 14 slides (27,598 tiles) from TRG0, 70 slides from TRG1/2 (98,102 tiles), and 28 slides (41,219 tiles) from TRG3. We included multiple slides per patient, a patient had one or two slides. After a series of data augmentation techniques, the augmented training set increased to three times the original training set. The test set was seven slides (15,506 tiles) from TRG0, seven slide (13,407 tiles) from TRG1/2, and six slides (10,263 tiles) from TRG3. For feature extracting, we utilized a deep residual network (Resnet101) without the final softmax classification layer and average pooling layer, pretrained with ImageNet. After the 2048-dimension feature vector were extracted, multiple full connected layers were used to non-linear mapping. At the end, a classification layer was added. In the training process, the tiles with 512 × 512 pixels were re-sized to 256 × 256 pixels and then input into the network. Among them, the patient's treatment responses were considered as the target label. We optimized the response classifier using a five-fold cross-validation approach, which involves randomly dividing the training set into five balanced subsets. Based on TRG classes, each fold was divided into 4:1 from different TRG class to extract training samples and validation samples. The hyperparameters include a batch size of 32, an initial learning rate of 0.001, and total epochs of 20 (decreasing learning rate by 0.1 every five epochs). In the post-processing stage, the average score of the tumor tiles of each WSI was taken as the prediction probability of the WSIs, which was regarded as the criterion to judge the efficacy of NAT. The response classifier structure was implemented in Python, using Pytorch backends (Python 3.9.7, Pytorch 1.11.0, SKLern 0.24.2). We used the same network structure, batch size, and learning rate to train and verify the model at 10× magnification. The initialized parameters with pre-training weights in ImageNet.

To analyze features used by the classifier to make its decision, we performed Smooth Grad-CAM (class activation mapping) ([Omeiza et al., 2019](#)) to identify regions within each

tile that the neural network uses to generate predictions. Then, we used cell detection and classification functions in Qupath to identify three types of cells: tumor cells, tumor-infiltrating lymphocytes (TILs), and stroma (*Acs et al., 2019*). TILs% calculated as follows:

$$TILs \% = \frac{TILs}{TILs + Tumor\ cells} \times 100\%.$$

Statistical analysis

We evaluated models at the tile and slide levels. The neural network yielded a probability value for every class of interest in each tile. Then we averaged the probabilities of each tile from the patients' slides to assign a final probability to each patient. The probabilities calculated by the neural network were used to generate the area under the curve (AUC) for the receiver operating characteristic (ROC) curve. Using macro-AUC and micro-AUC as metrics, we compared the performance of the models at different magnifications. We also used Precision, Sensitivity, Specificity, and the multi-class Matthews' correlation coefficients (MCC) to assess the performance of models. The statistical analyses were performed using SKLern 0.24.2 in Python (scikit-learn.).

RESULT

Workflow for predicting NAT response with a multi-class classification algorithm

A total of 94 patients were included in this study based on the inclusion/ exclusion criteria. [Table 1](#) summarizes the characteristics of TRGs of the cohort. The workflow of our studies is shown in [Fig. 1](#). Firstly, we collected and preprocessed the HE images from the patients, including scanning, delineating and segmenting the ROIs. In this step, a total of 166, 919 tiles were obtained. Then we normalized the color of tiles by the SPCN (*Vahadane et al., 2016*). Secondly, we input the tiles into the neural network in the form of tiles at 10× and 40× magnification, and extract the features through ResNet-101 to deduce the probabilities and predict the treatment responses: TRG0, TRG1/2, or TRG3. Finally, we constructed a visual heat map of tiles using CAM to explore the medical interpretability of the model.

Model performance evaluation under multi-class classification tasks on different scale

To test the model performance on different fields of view, HE tiles were preprocessed with two magnifications, including that of 40× and 10×. We first developed the response classifier at 40× magnification. At the tile level, the macro-AUC of the three-class classification model was 0.82, and the micro-AUC was 0.86 on the cross-validation of the training set, and the macro-AUC of 0.72 and the micro-AUC of 0.72 on the test set, as shown in [Figs. 2A](#) and [2E](#). The model revealed a stable performance with the macro-AUC of 0.97 and the micro-AUC of 0.96 at the slide level, as shown in [Fig. 2B](#). In order to compare the performance on different scales, with this dataset we trained and tested the classifier at 10× magnification. The AUCs of the tile and slide levels are shown in [Figs. 2C](#) and [2D](#). The macro-AUC and micro-AUC were 0.75, 0.78 and 0.79, 0.86, respectively. And the result

Table 1 Characteristics of the patients.

characteristics	TRG0	TRG1/2	TRG3	p value
Gender (%)				0.779
Male	12(63.2%)	32(62.7%)	17(70.8%)	
Female	7(36.8%)	19(37.3%)	7(29.2%)	
Age, mean \pm SD	52.74 \pm 10.76	56.65 \pm 11.36	56.50 \pm 13.23	0.443
Tumor location				0.205
Upper	2(15.4%)	3(6.1%)	4(18.2%)	
Middle	4(30.8%)	20(40.8%)	16(54.5%)	
Lower	7(53.8%)	26(53.1%)	6(27.3%)	
Pretreatment T stage (%)				0.083
T1	0(0%)	0(0%)	0(0%)	
T2	0(0%)	0(0%)	2(8.7%)	
T3	4(21.1%)	19(38.8%)	6(26.1%)	
T4	15(78.9%)	30(61.2%)	15(65.2%)	
Pretreatment N stage (%)				0.328
N0	1(5.3%)	3(6.0%)	4(16.7%)	
N1-N2	13(68.4%)	38(76.0%)	13(54.2%)	
Nx	5(26.3%)	9(18.0%)	7(29.2%)	
Pretreatment CEA				0.448
\leq 5 ng/ml	13(72.2%)	30(58.8%)	17(70.8%)	
$>$ 5 ng/ml	5(27.8%)	21(41.2%)	7(29.2%)	
Pretreatment CA199				0.879
\leq 34 U/ml	14(77.8%)	37(74.0%)	17(70.8%)	
$>$ 34 U/ml	4(22.2%)	13(26.0%)	7(29.2%)	
Neoadjuvant chemotherapy regime				0.856
Based on Capecitabine	16 (84.2%)	39 (76.5%)	19 (79.2%)	
Based on 5-FU	3 (15.8%)	11 (21.6%)	4 (16.7%)	
Cycles	4.39 \pm 1.82	3.55 \pm 2.27	3.48 \pm 2.87	0.383
Underwent neoadjuvant radiotherapy				0.009
Yes	18 (94.7%)	40 (78.4%)	12 (50.0%)	
No	1 (5.3%)	11 (21.6%)	11(45.8%)	

Notes.

CEA, carcinoembryonic antigen; CA199, carbohydrate antigen 19-9; SD, standard deviation.

of the test set and the variability on the 5-folds cross validation are shown in [Figs. 2E–2H](#), [Tables 2](#), and [3](#).

After comparing the overall performance of models, we determined the specific classification performance of the model for each group shown in [Fig. 3](#). Specifically, the performance of the model at the $40\times$ magnification is better than at the $10\times$ magnification, which indicates that the local cellular features extracted from the model were more conducive to the differentiation of curative effects than the global tissue features. The detailed AUCs under tile and slide level are provided in [Table 4](#).

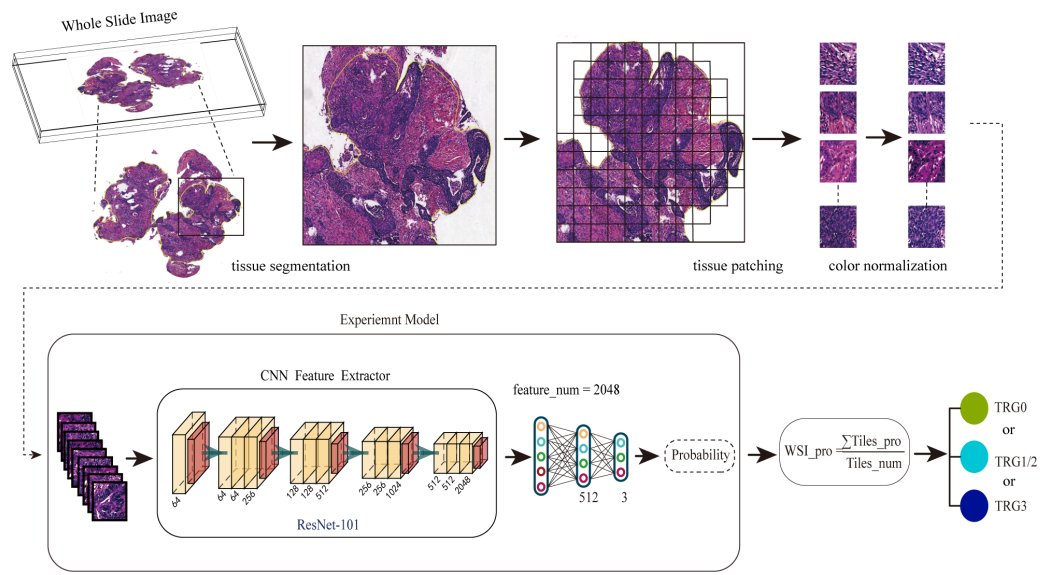


Figure 1 The workflow for predicting the NAT response using the multi-class classification algorithm. Firstly, we preprocessed the HE images, including scanning the images, delineating and segmenting the tumor tissues, and normalizing the color of tiles. Secondly, we input pathological images into the neural network in the form of tiles to deduce the probabilities. Finally, the WSIs were divided into different treatment responses by predicting probabilities.

Full-size DOI: [10.7717/peerj.15408/fig-1](https://doi.org/10.7717/peerj.15408/fig-1)

Medical interpretability analysis

To explore the relationship between histopathological features and the efficacy of NAT, we developed a post-hoc explanation of a trained model at the 40× and 10× magnifications, based on the CAM algorithm. The model generated a heat map for each tile, and the areas with large contribution from the model are displayed in red. As shown in Figs. 4A and 4B, overlaying the original tile images with the image of the CAM analyses revealed that tumor tissues including tumor cells and TILs seem to contribute the most in our model, while stroma contributes less to this model. Among them, TILs play an important role in the decision to classify responses, either with 40× or 10× magnifications. Specifically, tiles predicted with TRG0 appeared to be denser in the number of TILs than that with TRG1/2 or TRG3. As shown in Fig. 4E, it was observed that the proportion of TILs was significantly associated with the treatment response. The median TIL % in the TRG0 group was significantly higher than that in the other two groups (the TRG0 group was 35.1%, the TRG1/2 group 24.7%, and TRG3 group 16.9%).

Correlation analysis of TRG prediction

As shown in Table 5, at the 40× magnification, the MCC in the prediction of TRG0,1/2 and 3 at the tile level were 0.499, 0.501 and 0.425, respectively, while the MCC in the prediction of TRG0, 1/2 and 3 at the slide level were 0.617, 0.734, and 0.683, respectively. At the 10× magnification, the MCC in the prediction of TRG0,1/2 and 3 at the tile level were 0.623, 0.233 and 0.221, respectively, while the MCC in the prediction of TRG0, 1/2 and 3 at the slide level were 0.690, 0.316, and 0.298, respectively. According to the MCC

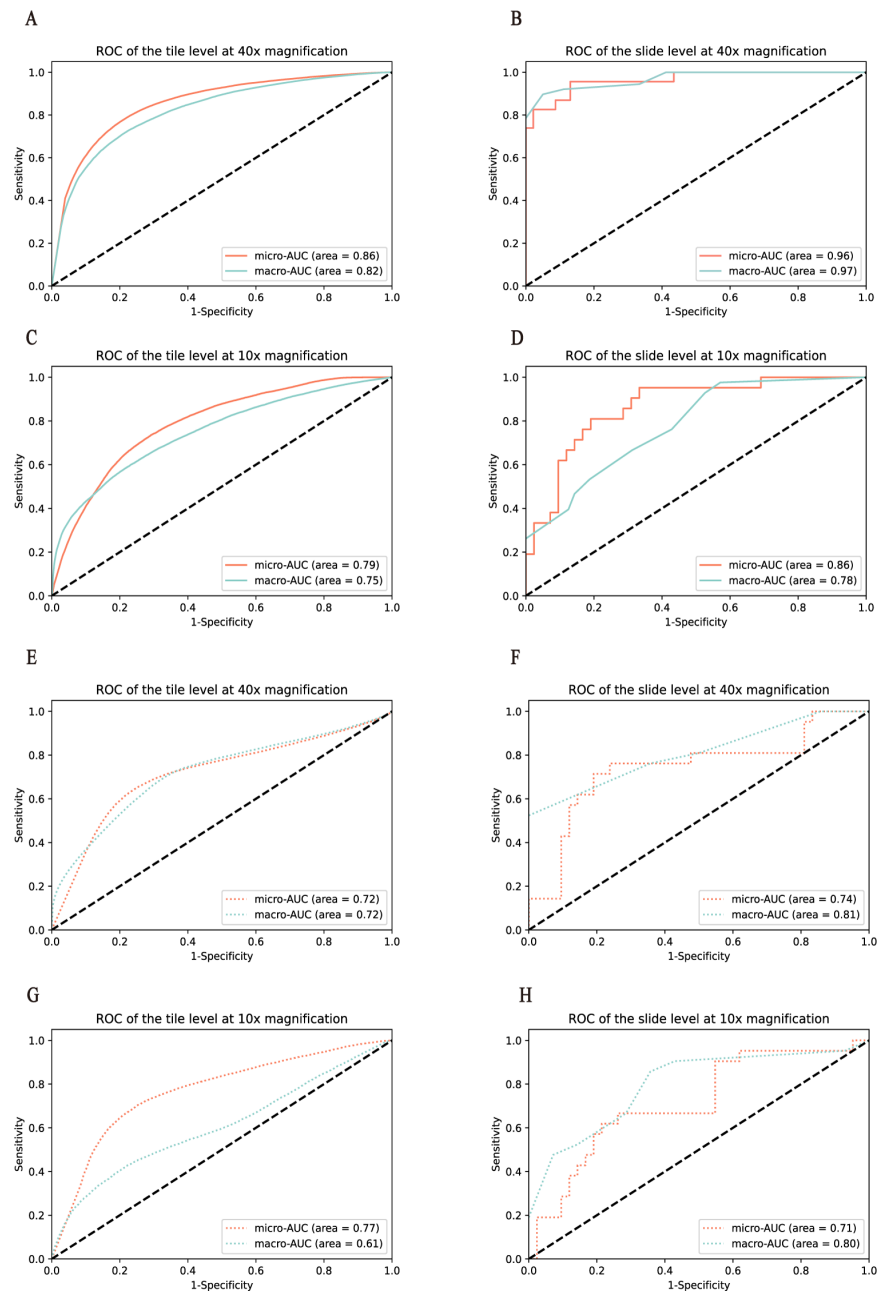


Figure 2 Performance of the model was measured in terms of the AUC of the ROC curves. Models were validated and tested on two levels. (A–D) represent the results of the model in the training set, and (E–H) represent the results of the model in the test set. Among them, (A, C, E, and G) represent the performance comparison between 40 \times and 10 \times magnification of the model at the tile level. (B, D, F, and H) represent the performance comparison between 40 \times and 10 \times magnification of the model at the slide level.

Full-size  DOI: [10.7717/peerj.15408/fig-2](https://doi.org/10.7717/peerj.15408/fig-2)

results, TRG1/2 and TRG3 were weakly differentiated at low magnification. We also found that the proportion of TILs in TRG1/2 and TRG3 was lower than that in TRG0, and this microstructure could not be clearly displayed at low magnification if the proportion was

Table 2 Macro-AUC and Micro-AUC for the training set and test set at 40× magnification.

	Slide level		Tile level	
	Macro-AUC	Micro-AUC	Macro-AUC	Micro-AUC
Validation 1	0.97	0.96	0.82	0.86
Validation 2	0.76	0.82	0.57	0.67
Validation 3	0.78	0.84	0.68	0.77
Validation 4	0.90	0.93	0.70	0.75
Validation 5	0.94	0.94	0.93	0.94
Test	0.81	0.74	0.72	0.72

Table 3 Macro-AUC and Micro-AUC for the training set and test set at 10× magnification.

	Slide level		Tile level	
	Macro-AUC	Micro-AUC	Macro-AUC	Micro-AUC
Validation 1	0.80	0.87	0.71	0.77
Validation 2	0.61	0.74	0.68	0.82
Validation 3	0.73	0.81	0.65	0.77
Validation 4	0.89	0.87	0.64	0.76
Validation 5	0.78	0.86	0.75	0.79
Test	0.80	0.71	0.61	0.77

low. Therefore, this may be the reason for the low MCC in pathological images with low magnification. Meanwhile, according to the MCC analysis results, we found that TRG0 had better predictability at low magnification, while TRG1/2 and TRG3 had relatively better predictability at high magnification. Therefore, combination of low magnification and high magnification pathological images to construct a multi-scale joint prediction model might improve the accuracy and robustness of TRG classification prediction for NAT.

DISCUSSION

For locally advanced rectal cancer, NAT has become a standard treatment method that can improve the quality of life and overall survival. Despite the benefit of the treatment modality, about 40% of patients are not benefited from it. For this reason, identification of biomarkers (biological or multiomics signatures) or image markers (from radiological and/or pathological image features) for NAT responses is of clinical significance. Risk stratification for NAT represents the opportunity to change the chemo-radiotherapy regimens or switch to other treatments as early as possible. However, most studies toward this goal are limited to the prediction power to identifying pCR patients. Here we present a new multi-class DL algorithm that can identify three categories of the NAT responses: very sensitive (TRG0), moderately sensitive (TRG1/2) or resistant (TRG3). The performance of the algorithm is demonstrated in the two magnifications.

Over the last few years, various ML and DL algorithms have been applied to pathological image classification tasks, which allow disease identification and risk stratification. Studies have shown that pathological images can provide sub-visual information about molecular

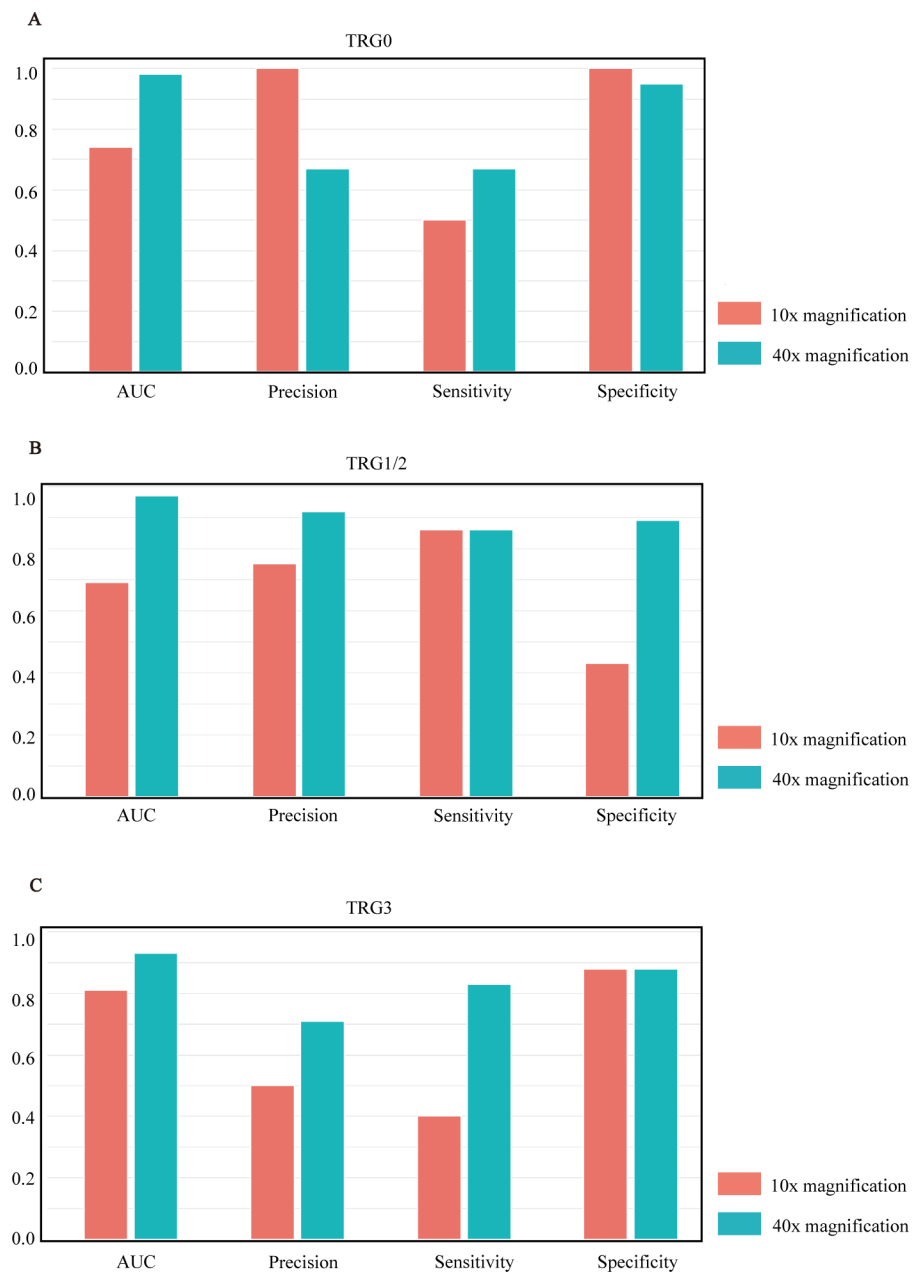


Figure 3 Model performance histograms of the slides at different magnifications. (A–C) respectively represent the comparison of AUC, precision, sensitivity, and specificity of the model against TRG0, TRG1/2, and TRG3 at 10× and 40× magnification.

Full-size  DOI: [10.7717/peerj.15408/fig-3](https://doi.org/10.7717/peerj.15408/fig-3)

and genomic characteristics to supplement and enhance the predictive ability. There are three studies (*Zhang et al., 2020*; *Shao et al., 2020*; *Feng et al., 2022*) using pathological HE images to predict the curative effect of NAT for rectal cancer. Although these studies have achieved powerful prediction powers, it is only for identification of pCR vs non-pCR. These programs, as well as other ML-based programs for radiological images, lack the power of

Table 4 AUC for three-class classification model at 40× magnification and 10× magnification under the slide level and tile level.

category		At 40× magnification	At 10× magnification
TRG0	Slide level	0.98	0.74
	Tile level	0.87	0.92
TRG1/2	Slide level	0.97	0.69
	Tile level	0.83	0.67
TRG3	Slide level	0.93	0.81
	Tile level	0.77	0.67

Notes.

AUC, area under the curve.

Table 5 The multi-class Matthews' correlation coefficients in the prediction of TRG 0,1/2 and 3 at different magnifications.

category		At 40× magnification	At 10× magnification
TRG0	Slide level	0.617	0.690
	Tile level	0.499	0.623
TRG1/2	Slide level	0.734	0.316
	Tile level	0.501	0.233
TRG3	Slide level	0.683	0.298
	Tile level	0.425	0.221

identifying TRG3 patients, the most important group of patients that will not benefit from NAT at all. Therefore, our DL-based algorithm has more clinical significance.

The image of CAM analyses revealed that TILs seem to contribute to the most in our model. TILs are considered to be important anti-tumor cells ([Rosenberg, Spiess & Lafreniere, 1986](#)). As large numbers of lymphocytes infiltrating into the tumor, they contribute to the response to radiotherapy and chemotherapy. It is known that the number of TILs have positive significance for the tumor response and survival outcome. Accumulation of TILs at tumor sites is taken as a positive prognostic factor in rectal cancer ([Zinovkin et al., 2021](#)). Therefore, this pathological feature is statistically related to the great response, which provides an explanation for some features extracted by our model when determining the treatment response.

There are limitations of our study: Firstly, data used to build the model is from a single center. Multi-center samples for cross-center validation are needed, in order to enhance the generalization ability of the model. Secondly, the model was based on single scale images. In addition to experiments in the single scale, fusing with 10× and 40× images to jointly build a multi-scale model might improve the performance. Finally, we still need to investigate the relationship between efficacy-related features extracted from pathological images and overall survival outcomes for the disease.

In conclusion, the DL-based multi-class algorithm we established provides a superior prediction power for rectal cancer NAT.

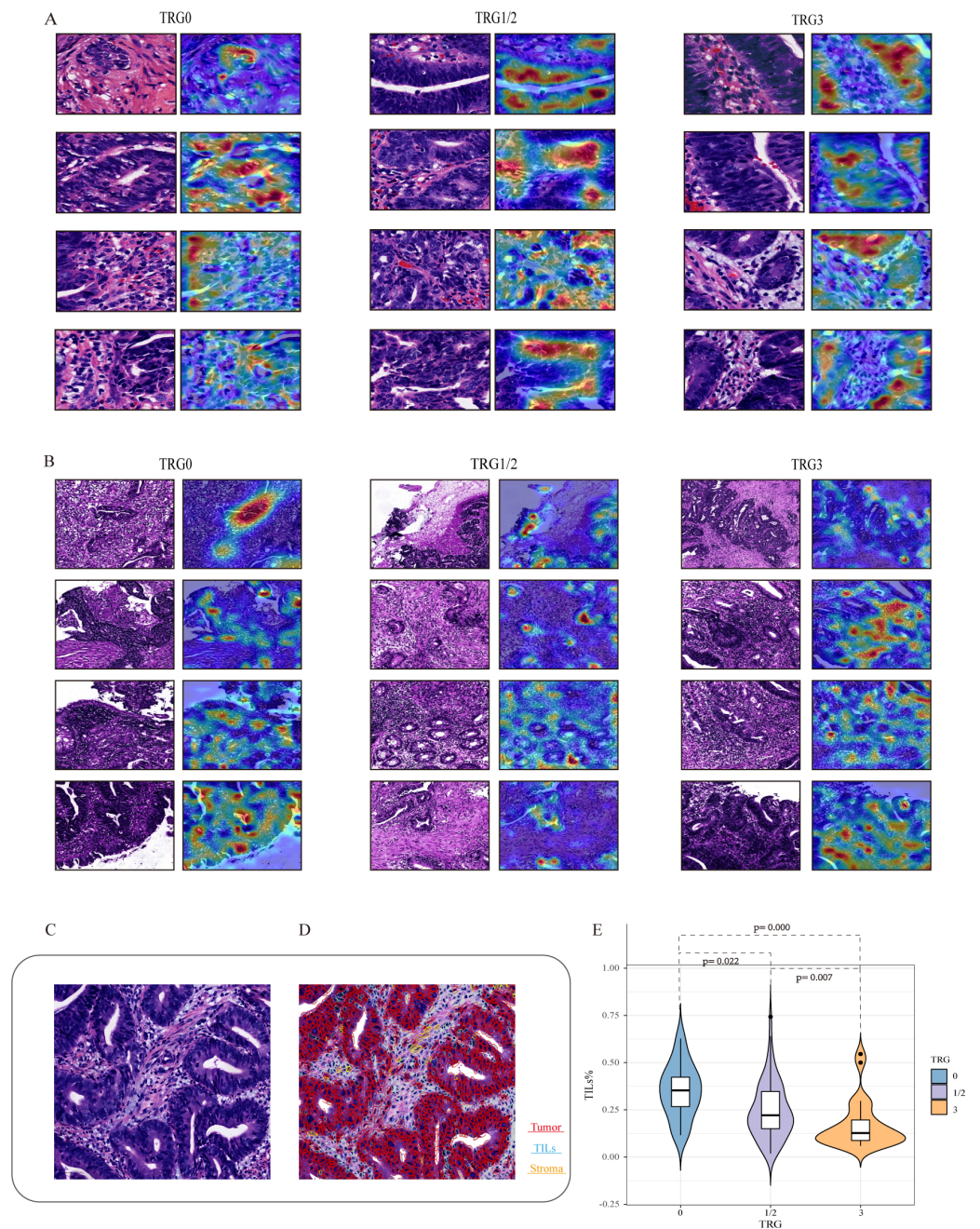


Figure 4 Visualization of the important regions in HE images at 40 \times and 10 \times magnifications for the prediction of NAT. Based on the CAM algorithm, areas with large contribution from the model are displayed in red, and (A and B) represent the visual expression of different treatment responses at 40 \times and 10 \times magnifications. We visualize important local tumor areas in WSIs to explore differences between the efficacy groups. (C and D) represent the representative HE images and the cell quantitative analysis mask. Using the Qupath, segmentation shows red indicates tumor cells, blue indicates TILs, and yellow indicates stromal cells. (E) shows the proportion of TILs was significantly associated with the treatment response. Full-size [DOI: 10.7717/peerj.15408/fig-4](https://doi.org/10.7717/peerj.15408/fig-4)

ACKNOWLEDGEMENTS

We thank all members of the Xu laboratory for helpful discussions.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the National Natural Science Foundation of China [61906022], Chongqing Natural Science Foundation cstc2020jcyj-msxmX0482, Chongqing University Research Fund 2021CDJXKJC004, and Chongqing Medical Scientific Research project (Joint project of Chongqing Health Commission and Science and Technology Bureau) 2020MSXM088. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

The National Natural Science Foundation of China: 61906022.

Chongqing Natural Science Foundation: cstc2020jcyj-msxmX0482.

Chongqing University Research Fund: 2021CDJXKJC004.

Chongqing Medical Scientific Research project (Joint project of Chongqing Health Commission and Science and Technology Bureau): 2020MSXM088.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yihan Wu performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Xiaohua Liu performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Fang Liu performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Yi Li analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Xiaomin Xiong performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Hao Sun conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Bo Lin conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Yu Li conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Bo Xu conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

This study was approved by the Chongqing University Cancer Hospital of Medicine Institutional Review Board (Ethical Application Ref: CZLS2022185-A).

Data Availability

The following information was supplied regarding data availability:

The code and raw measurements are available in the [Supplementary Files](#). The raw data show the probability values of slides and tiles judged to be in each category at different magnifications.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.15408#supplemental-information>.

REFERENCES

- Acs B, Ahmed FS, Gupta S, Wong PF, Gartrell RD, Sarin Pradhan J, Rizk EM, Gould Rothberg B, Saenger YM, Rimm DL. 2019. An open source automated tumor infiltrating lymphocyte algorithm for prognosis in melanoma. *Nature Communications* 10(1):5440 DOI 10.1038/s41467-019-13043-2.
- Belluco C, De Paoli A, Canzonieri V, Sigon R, Fornasarig M, Buonadonna A, Boz G, Innocente R, Perin T, Cossaro M, Polesel J, De Marchi F. 2011. Long-term outcome of patients with complete pathologic response after neoadjuvant chemoradiation for cT3 rectal cancer: implications for local excision surgical strategies. *Annals of Surgical Oncology* 18(13):3686–3693 DOI 10.1245/s10434-011-1822-0.
- Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, Cohen S, Cooper HS, Deming D, Garrido-Laguna I, Grem JL, Gunn A, Hoffe S, Hubbard J, Hunt S, Kirilcuk N, Krishnamurthi S, Messersmith WA, Meyerhardt J, Miller ED, Mulcahy MF, Nurkin S, Overman MJ, Parikh A, Patel H, Pedersen K, Saltz L, Schneider C, Shibata D, Skibber JM, Sofocleous CT, Stoffel EM, Stotsky-Himelfarb E, Willett CG, Johnson-Chilla A, Gurski LA. 2020. NCCN Guidelines insights: rectal cancer, version 6. *Journal of the National Comprehensive Cancer Network* 18(7):806–815 DOI 10.6004/jnccn.2020.0032.
- Bilal M, Raza SEA, Azam A, Graham S, Ilyas M, Cree IA, Minhas F, Rajpoot NM. 2021. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *The Lancet Digital Health* 3(12):e763–e772.
- de Campos-Lobato LF, Stocchi L, da Luz Moreira A, Geisler D, Dietz DW, Lavery IC, Fazio VW, Kalady MF. 2011. Pathologic complete response after neoadjuvant treatment for rectal cancer decreases distant recurrence and could eradicate local recurrence. *Annals of Surgical Oncology* 18(6):1590–1598 DOI 10.1245/s10434-010-1506-1.

- National Health Commission of the People's Republic of China.** 2020. Chinese protocol of diagnosis and treatment of colorectal cancer (2020 edition). *Chinese Journal of Digestive Surgery* **19**(6):563–588.
- Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, Moreira AL, Razavian N, Tsirigos A.** 2018. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine* **24**(10):1559–1567 DOI [10.1038/s41591-018-0177-5](https://doi.org/10.1038/s41591-018-0177-5).
- Courtiol P, Maussion C, Moarii M, Pronier E, Pilcer S, Sefta M, Manceron P, Toldo S, Zaslavskiy M, Stang NLe, Girard N, Elemento O, Nicholson AG, Blay JY, Galateau-Salle F, Wainrib G, Clozel T.** 2019. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature Medicine* **25**(10):1519–1525 DOI [10.1038/s41591-019-0583-3](https://doi.org/10.1038/s41591-019-0583-3).
- Das P, Skibber JM, Rodriguez-Bigas MA, Feig BW, Chang GJ, Wolff RA, Eng C, Krishnan S, Janjan NA, Crane CH.** 2007. Predictors of tumor response and downstaging in patients who receive preoperative chemoradiation for rectal cancer. *Cancer* **109**(9):1750–1755 DOI [10.1002/cncr.22625](https://doi.org/10.1002/cncr.22625).
- Feng L, Liu Z, Li C, Li Z, Lou X, Shao L, Wang Y, Huang Y, Chen H, Pang X, Liu S, He F, Zheng J, Meng X, Xie P, Yang G, Ding Y, Wei M, Yun J, Hung M-C, Zhou W, Wahl DR, Lan P, Tian J, Wan X.** 2022. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *The Lancet Digital Health* **4**(1):e8–e17.
- Huh JW, Lee JH, Kim HR.** 2014. Pretreatment expression of 13 molecular markers as a predictor of tumor responses after neoadjuvant chemoradiation in rectal cancer. *Annals of Surgery* **259**(3):508–515 DOI [10.1097/SLA.0b013e31829b3916](https://doi.org/10.1097/SLA.0b013e31829b3916).
- Jepsen DNM, Høeg H, Thagaard J, Walbech JS, Gögenur I, Fiehn AK.** 2022. Development of a method for digital assessment of tumor regression grade in patients with rectal cancer following neoadjuvant therapy. *Journal of Pathology Informatics* **13**:100152 DOI [10.1016/j.jpi.2022.100152](https://doi.org/10.1016/j.jpi.2022.100152).
- Karagkounis G, Thai L, Mace AG, Wiland H, Pai RK, Steele SR, Church JM, Kalady MF.** 2019. Prognostic implications of pathological response to neoadjuvant chemoradiation in pathologic stage III rectal cancer. *Annals of Surgery* **269**(6):1117–1123 DOI [10.1097/SLA.0000000000002719](https://doi.org/10.1097/SLA.0000000000002719).
- Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis CA, Gaiser T, Marx A, Valous NA, Ferber D, Jansen L, Reyes-Aldasoro CC, Zornig I, Jager D, Brenner H, Chang-Claude J, Hoffmeister M, Halama N.** 2019. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLOS Medicine* **16**(1):e1002730 DOI [10.1371/journal.pmed.1002730](https://doi.org/10.1371/journal.pmed.1002730).
- Kim SH, Chang HJ, Kim DY, Park JW, Baek JY, Kim SY, Park SC, Oh JH, Yu A, Nam BH.** 2016. What is the ideal tumor regression grading system in rectal cancer patients after preoperative chemoradiotherapy? *Cancer Research and Treatment* **48**(3):998–1009 DOI [10.4143/crt.2015.254](https://doi.org/10.4143/crt.2015.254).
- Kurata Y, Hayano K, Ohira G, Imanishi S, Tochigi T, Isozaki T, Aoyagi T, Matsubara H.** 2021. Computed tomography-derived biomarker for predicting the treatment

- response to neoadjuvant chemoradiotherapy of rectal cancer. *International Journal of Clinical Oncology* 26(12):2246–2254 DOI 10.1007/s10147-021-02027-2.
- Lin B, Tan Z, Mo Y, Yang X, Liu Y, Xu B. 2022.** Intelligent oncology: the convergence of artificial intelligence and oncology. *Journal of the National Cancer Center* in press.
- Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, Wang S, Li XT, Tian J, Sun YS. 2017.** Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clinical Cancer Research* 23(23):7253–7262 DOI 10.1158/1078-0432.CCR-17-1038.
- Matsutani S, Shibutani M, Maeda K, Nagahara H, Fukuoka T, Nakao S, Hirakawa K, Ohira M. 2018.** Significance of tumor-infiltrating lymphocytes before and after neoadjuvant therapy for rectal cancer. *Cancer Science* 109(4):966–979 DOI 10.1111/cas.13542.
- Monson JRT, Arsalanizadeh R. 2017.** Surgery for patients with rectal cancer-time to listen to the patients and recognize reality. *JAMA Oncology* 3(7):887–888 DOI 10.1001/jamaoncol.2016.5380.
- Nicolas AM, Pesic M, Engel E, Ziegler PK, Diefenhardt M, Kennel KB, Buettner F, Conche C, Petrocelli V, Elwakeel E, Weigert A, Zinoveva A, Fleischmann M, Häupl B, Karakütük C, Bohnenberger H, Mosa MH, Kaderali L, Gaedcke J, Ghadimi M, Rödel F, Arkan MC, Oellerich T, Rödel C, Fokas E, Greten FR. 2022.** Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer. *Cancer Cell* 40(2):168–184.e13 DOI 10.1016/j.ccell.2022.01.004.
- Omeiza D, Speakman S, Cintas C, Weldermariam K. 2019.** Smooth Grad-CAM++: an enhanced inference level visualization technique for deep convolutional neural network models. In: *Intelligent systems conference 2019*.
- Rosenberg SA, Spiess P, Lafreniere R. 1986.** A new approach to the adoptive immunotherapy of cancer with tumor-infiltrating lymphocytes. *Science* 233(4770):1318–1321 DOI 10.1126/science.3489291.
- Shao L, Liu Z, Feng L, Lou X, Li Z, Zhang XY, Wan X, Zhou X, Sun K, Zhang DF, Wu L, Yang G, Sun YS, Xu R, Fan X, Tian J. 2020.** Multiparametric MRI and whole slide image-based pretreatment prediction of pathological response to neoadjuvant chemoradiotherapy in rectal cancer: a multicenter radiopathomic study. *Annals of Surgical Oncology* 27(11):4296–4306 DOI 10.1245/s10434-020-08659-4.
- She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, Jiang G, Liu H, Xie D, Cao N, Ren Y, Chen C. 2020.** Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA Network Open* 3(6):e205842 DOI 10.1001/jamanetworkopen.2020.5842.
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. 2021.** Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 71(3):209–249.
- Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, Navab N. 2016.** Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Transactions on Medical Imaging* 35(8):1962–1971 DOI 10.1109/TMI.2016.2529665.

- Wallin U, Rothenberger D, Lowry A, Luepker R, Mellgren A. 2013.** CEA—a predictor for pathologic complete response after neoadjuvant therapy for rectal cancer. *Diseases of the Colon and Rectum* **56(7)**:859–868 DOI [10.1097/DCR.0b013e31828e5a72](https://doi.org/10.1097/DCR.0b013e31828e5a72).
- Wang H, Ji D, Tian H, Gao Z, Song C, Jia J, Cui X, Zhong L, Shen J, Gu J. 2022.** Predictive value of proteomic markers for advanced rectal cancer with neoadjuvant chemoradiotherapy. *BMC Cancer* **22(1)**:868 DOI [10.1186/s12885-022-09960-z](https://doi.org/10.1186/s12885-022-09960-z).
- Yamashita R, Long J, Longacre T, Peng L, Berry G, Martin B, Higgins J, Rubin DL, Shen J. 2021.** Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *The Lancet Oncology* **22(1)**:132–141 DOI [10.1016/S1470-2045\(20\)30535-0](https://doi.org/10.1016/S1470-2045(20)30535-0).
- Zhang F, Yao S, Li Z, Liang C, Zhao K, Huang Y, Gao Y, Qu J, Li Z, Liu Z. 2020.** Predicting treatment response to neoadjuvant chemoradiotherapy in local advanced rectal cancer by biopsy digital pathology image features. *Clinical and Translational Medicine* **10(2)**:e110.
- Zhou D, Tian F, Tian X, Sun L, Huang X, Zhao F, Zhou N, Chen Z, Zhang Q, Yang M, Yang Y, Guo X, Li Z, Liu J, Wang J, Wang B, Zhang G, Sun B, Zhang W, Kong D, Chen K, Li X. 2020.** Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. *Nature Communications* **11(1)**:2961 DOI [10.1038/s41467-020-16777-6](https://doi.org/10.1038/s41467-020-16777-6).
- Zhu M, Li X, Cheng X, Yi X, Ye F, Li X, Hu Z, Zhang L, Nie J, Li X. 2022.** Association of the tissue infiltrated and peripheral blood immune cell subsets with response to radiotherapy for rectal cancer. *BMC Medical Genomics* **15(Suppl 2)**:107 DOI [10.1186/s12920-022-01252-6](https://doi.org/10.1186/s12920-022-01252-6).
- Zinovkin DA, Kose SY, Nadyrov EA, Achinovich SL, Los DM, Gavrilenko TE, Gavrilenko DI, Yuzugulen J, Pranjol MZI. 2021.** Potential role of tumor-infiltrating T-, B-lymphocytes, tumor-associated macrophages and IgA-secreting plasma cells in long-term survival in the rectal adenocarcinoma patients. *Life Sciences* **286**:120052 DOI [10.1016/j.lfs.2021.120052](https://doi.org/10.1016/j.lfs.2021.120052).