Retrospective Cohort Study

INTERNATIONAL JOURNAL OF SURGERY

OPEN

A computed tomography-based multitask deep learning model for predicting tumour stroma ratio and treatment outcomes in patients with colorectal cancer: a multicentre cohort study

Yanfen Cui, PhD[a,b,c,e], Ke Zhao, MD[a,b,c], Xiaochun Meng, PhD[d], Yun Mao, PhD[f], Chu Han, PhD[a,b,c], Zhenwei Shi, PhD[a,b,c], Xiaotang Yang, PhD[e,*], Tong Tong, PhD[g,*], Lei Wu, PhD[a,b,c,*], Zaiyi Liu, PhD[a,b,c,*]

Background: Tumour-stroma interactions, as indicated by tumour-stroma ratio (TSR), offer valuable prognostic stratification information. Current histological assessment of TSR is limited by tissue accessibility and spatial heterogeneity. The authors aimed to develop a multitask deep learning (MDL) model to noninvasively predict TSR and prognosis in colorectal cancer (CRC).

Materials and methods: In this retrospective study including 2268 patients with resected CRC recruited from four centres, the authors developed an MDL model using preoperative computed tomography (CT) images for the simultaneous prediction of TSR and overall survival. Patients in the training cohort ($n = 956$) and internal validation cohort (IVC, $n = 240$) were randomly selected from centre I. Patients in the external validation cohort 1 (EVC1, $n = 509$), EVC2 ($n = 203$), and EVC3 ($n = 360$) were recruited from other three centres. Model performance was evaluated with respect to discrimination and calibration. Furthermore, the authors evaluated whether the model could predict the benefit from adjuvant chemotherapy.

Results: The MDL model demonstrated strong TSR discrimination, yielding areas under the receiver operating curves (AUCs) of 0.855 (95% CI, 0.800–0.910), 0.838 (95% CI, 0.802–0.874), and 0.857 (95% CI, 0.804–0.909) in the three validation cohorts, respectively. The MDL model was also able to predict overall survival and disease-free survival across all cohorts. In multivariable Cox analysis, the MDL score (MDLS) remained an independent prognostic factor after adjusting for clinicopathological variables (all $P < 0.05$). For stage II and stage III disease, patients with a high MDLS benefited from adjuvant chemotherapy [hazard ratio (HR) 0.391 (95% CI, 0.230–0.666), $P = 0.0003$; HR = 0.467 (95% CI, 0.331–0.659), $P < 0.0001$, respectively], whereas those with a low MDLS did not.

Conclusion: The multitask DL model based on preoperative CT images effectively predicted TSR status and survival in CRC patients, offering valuable guidance for personalized treatment. Prospective studies are needed to confirm its potential to select patients who might benefit from chemotherapy.

Keywords: Colorectal cancer, deep learning, survival, tumour-stroma ratio

[a]Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, [b]Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, [c]Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, [d]Department of Radiology, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, [e]Department of Radiology, Shanxi Province Cancer Hospital/ Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/Cancer Hospital Affiliated to Shanxi Medical University; Taiyuan, [f]Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing and [g]Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China

Y.C., K.Z., X.M. and Y.M. contributed equally to this work.

X.Y., T.T., L.W. and Z.L. are the co-corresponding authors of this study.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

**Corresponding authors. Address: Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, 106 Zhongshan Er Road, Guangzhou, 510080, China. Tel.: +861 392 2369 345. E-mail: liuzaiyi@gdph.org.cn (Z. Liu), and Tel.: +861 343 0275 861. E-mail: wulei@gdph.org.cn (L. Wu); Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai 200032, China. Tel.: +861 801 7312 912. E-mail: t983352@126.com (T. Tong); Department of Radiology, Shanxi Province Cancer Hospital/ Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/Cancer Hospital Affiliated to Shanxi Medical University; Taiyuan, 030013, China. Tel.: +861 593 5151 002. E-mail: yangxt210@126.com (X. Yang).*

International Journal of Surgery (2024) 110:2845–2854

Received 17 October 2023; Accepted 26 January 2024

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.lww.com/international-journal-of-surgery.

Published online 12 February 2024

http://dx.doi.org/10.1097/JS9.0000000000001161

Introduction

Colorectal cancer (CRC) ranks as the third most frequently diagnosed cancer and stands as the second leading cause of cancer-related deaths worldwide[1]. The established TNM staging system serve as the cornerstone for decision-making regarding treatment and forecasting outcomes in CRC[2]. However, due to the tumour heterogeneity, divergent prognosis was observed among patients even with same TNM stage[3]. Innovative biomarkers are urgently needed to enhance prognostic stratification and inform personalized treatment decisions.

Evolving evidence underscores the significance of the tumour microenvironment (TME), particularly the tumour-associated stroma, in driving aggressive behaviours, such as local and metastatic spread, and potentially influencing resistance to chemotherapy[4,5]. Amongst the stroma-related TME biomarkers, the tumour-stroma ratio (TSR), reflecting the interplay between tumour and stromal elements, has emerged as a stage-independent indicator for survival prognosis in patients with CRC[6,7]. Yet, existing histological evaluation of TSR relies on post-surgery tissue samples, which introduces sampling bias due to intratumor spatial heterogeneity and limits accessibility[8]. Moreover, visual estimation is prone to inconsistencies among pathologists, leading to discrepancies[9]. Consequently, a noninvasive approach for assessing TSR status becomes imperative, enabling impartial and longitudinal TME evaluations.

Radiographical imaging is routinely used clinically for staging and evaluating treatment response in CRC, encompassing a wealth of information about tumour phenotypes. Radiomics, an innovative strategy, converting medical images into high-throughput quantitative features, proposing an innovative path for noninvasive tumour and TME evaluation[10–12]. Several studies have sought to explore the association between specific radiomics features and the TME, such as tumour-infiltrating lymphocytes, immunoscore, and neutrophil-to-lymphocyte ratio (NLR), with moderate performance[13–15]. However, these classic radiomics approaches rely heavily on handcrafted feature engineering, and requires time-consuming meticulous tumour labelling, hindering practical clinical application[16].

Deep learning (DL), which automatically learn representation information directly from raw images, obviating manual feature engineering by domain experts, has garnered increasing attention[17,18]. Numerous investigations have substantiated the performance of DL in clinical diagnosis, prognosis prediction, and treatment options in many types of cancers, including CRC[19,20]. Early results have demonstrated the ability of DL to predict TME and treatment outcomes in gastric cancer patients[21,22]. Nevertheless, the relationship between DL and the TSR in TME of CRC remains uncertain. Notably, DL models are tailored to specific tasks, while multitask learning can share feature representations among related tasks, possibly mitigating overfitting and enhancing model generalization[23].

Thus, our goal is to develop a multitask deep learning model, harnessing preoperative computed tomography (CT) images, to simultaneously predict TSR status and overall survival in a large-scale multicenter CRC patient cohort. We also explored the model's ability to predict the benefit from adjuvant chemotherapy, as the secondary goal of this study.

HIGHLIGHTS

- We developed and validated a multitask deep learning model from pretreatment computed tomography images to noninvasively predict tumour-stroma ratio and prognosis in patients with colorectal cancer.
- The proposed multitask deep learning signature has promising performance in predicting adjuvant chemotherapy in stage II and stage III disease.

Materials and methods

Study design

An overview of the study is presented in Fig. 1. A total of 2268 consecutive patients with histologically confirmed CRC from four independent hospitals in China were included. For TSR and survival prediction, the training cohort (TC, $n = 956$) and internal validation cohort (IVC, $n = 240$) were retrospectively recruited at Center I from January 2014 to December 2016. Additionally, 509 patients from Center II and 203 patients from Center III in South China, were designated as external validation cohort 1 (EVC1) and EVC2, respectively. For prognostic prediction only, external validation cohort 3 (EVC3) comprising 360 patients from Center IV, was collected between June 2013 and December 2015. The detailed inclusion and exclusion criteria are outlined in Appendix E1, Supplemental Digital Content 1, http://links.lww.com/JS9/B923 and Figure S1, Supplemental Digital Content 1, http://links.lww.com/JS9/B923.

Baseline clinicopathological features, such as age, gender, tumour location, carcinoembryonic antigen (CEA) level, as well as pathological tumour and lymph node stages, according to the 8th AJCC TNM staging system[24], were gathered from medical records. The primary outcome measured was overall survival (OS), defined as the time from surgery to death from any cause. The secondary outcome, disease-free survival (DFS), was defined as the interval to either disease progression or death from any cause. Patients alive and disease-free at the last follow-up were censored. All patients were postoperatively followed every 3–6 months for the first 2 years, then every 6 months during the next 3 years, and then annually thereafter.

This study was approved by the institutional review boards of all participating hospitals, and informed consent was waived owing to the observational and retrospective nature of the study. This study was also approved by Chinese Clinical Trial (ChiCTR20000635734), and was reported in line with the STROCSS, Supplemental Digital Content 2, http://links.lww.com/JS9/B924 (Strengthening The Reporting of Cohort Studies in Surgery) criteria[25].

Automatic computation of the TSR

Different from traditional visual microscopic assessment, we developed a deep learning (DL) model for quantifying TSR based on histological whole-slide images (WSI)[26]. The TSR was computed as the area of stroma divided by the total area of stroma and tumour taken together, yielding the final TSR score (Figure S2, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). The TSR status was then categorized into two groups: stroma-high group ($> 50\%$) and stroma-low group ($\leq 50\%$), based on a pre-established threshold[27]. More information is

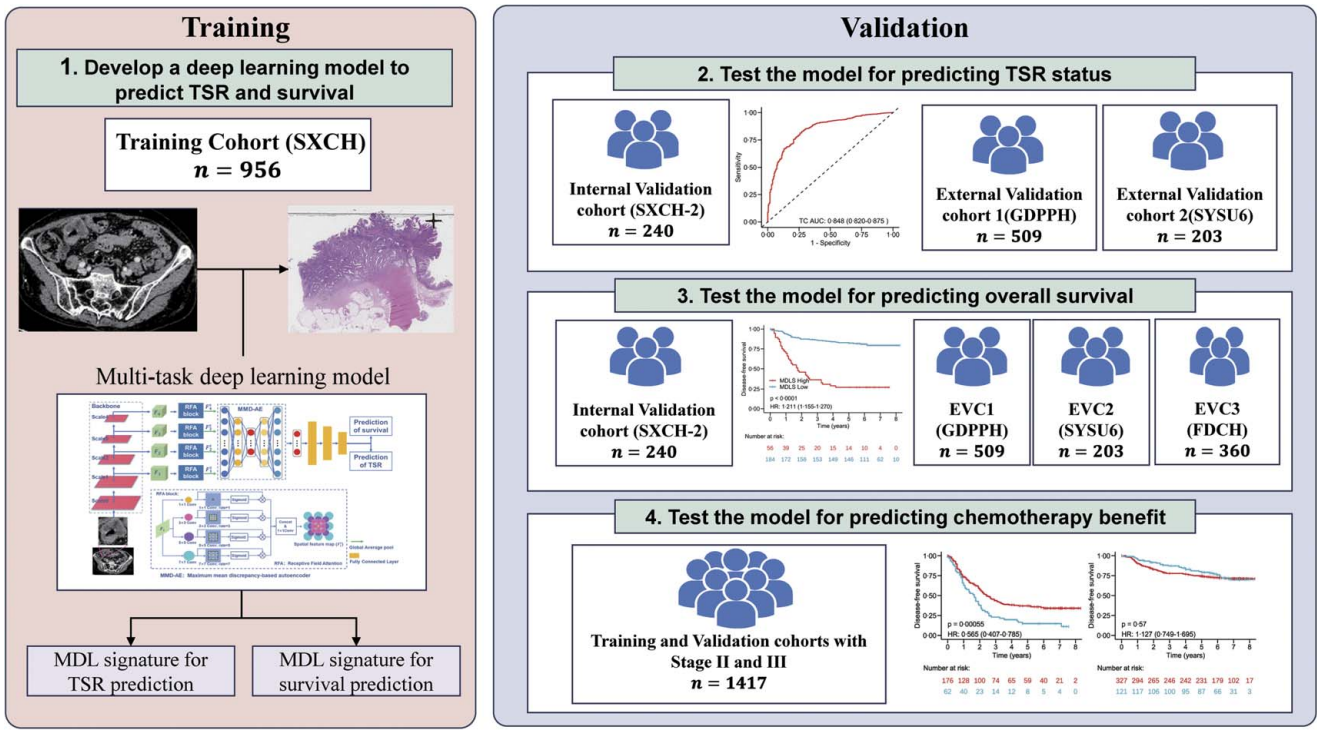

Figure 1. Study design for the training and validation of the multitask deep learning model based on computed tomography images to predict tumour-stroma ratio status and overall survival in patients with colorectal cancer. EVC1, external validation cohort 1; EVC2, external validation cohort 2; EVC3, external validation cohort 3; MDL, multitask deep learning; TSR, tumour-stroma ratio.

available in Appendix E2, Supplemental Digital Content 1, http://links.lww.com/JS9/B923.

Acquisition and analysis of CT images

CT image acquisition and preprocessing are detailed in Appendix E3, Supplemental Digital Content 1, http://links.lww.com/JS9/B923 and Table S2, Supplemental Digital Content 1, http://links.lww.com/JS9/B923. Two experienced radiologists, with 12 and 10 years of abdominal CT interpretation, respectively, delineated the tumour contours manually on the largest tumour section of the portal-venous phase CT images using the ITK-SNAP (version 3.8.0; http://www.itksnap.org). Notably, large vessels, adjacent organs, pericolonic fat, and air cavities were excluded.

Deep learning model development and model accuracy for TSR prediction

For the prediction of TSR status and survival in CRC patients based on CT images, we established a multitask deep learning (MDL) model (Fig. 1). Detail regarding model development and training process are in Appendix E4-5, Supplemental Digital Content 1, http://links.lww.com/JS9/B923. Gradient-weighted class activation mapping was utilized to visualize the MDL output and relevant regions of the CT images.

The diagnostic accuracy of the MDL model for TSR prediction was quantified by using the area under the receiver operating characteristic curve (AUC). Additionally, the corresponding metrics, including accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. This evaluation encompassed patients from the TC, IVC, EVC1, and EVC2, where TSR data were available.

Association with prognosis and benefit of chemotherapy

We evaluated the prognostic efficacy of the MDL model for predicting OS and DFS across all five independent cohorts. The optimal threshold for the predicted survival-score to categorize patients into high-risk or low-risk groups, was identified by using the maximally selected rank statistics method in the TC, and then applied to other validation cohorts. Stratified analyses were performed in subgroups defined by clinicopathological risk factors. Model performance was measured by Harrell's concordance index (C-index).

Furthermore, we developed an integrated model by combing the MDL signature with clinicopathological factors, for individualized assessment of OS and DFS. The net reclassification index (NRI) was calculated to quantify the relative improvement in prediction accuracy. The overall performance of these models was assessed with the prediction error curves and integrated Brier scores (IBS). Calibration curves were generated to compare the predicted survival probabilities with the actual probabilities for the outcome of interest. Furthermore, the ability of the MDL model to predict the benefit of adjuvant chemotherapy was evaluated among patients with stage II and III CRC.

Ablation analysis for multitask deep learning

To validate the efficacy of the proposed approach for simultaneous prediction of TSR status and OS, a comparative analysis is

conducted against both single-task TSR status prediction network and survival prediction network. Furthermore, to underscore the significance of incorporating multiscale features, a comparative evaluation is performed against a single-scale model.

Moreover, in an effort to validate the effectiveness of each individual module within the model, specific modules, including receptive field attention (RFA) and maximum mean discrepancy-based autoencoder (MMD-AE) modules, are intentionally removed. Finally, to further verify the effectiveness of our model, we compared it with other commonly employed models including fine-tuned multitask VGG16 and DenseNet121.

Statistical analysis

Continuous variables were analyzed with the student *t*-test or Mann–Whitney U test, while categorical data with the Pearsonχ^2 test or Fisher's exact test, as appropriate. Survival probabilities were evaluated by Kaplan–Meier survival analysis and log-rank test. Univariate and multivariate Cox regression analyses were conducted to select candidate predictors of survival. Interaction between the MDL signature and adjuvant chemotherapy was evaluated by the Cox model. All statistical analyses were carried out using R (version 3.6.2, https://www.r-project.org, Appendix E6, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). *P* values less than 0.05 were regarded as statistically significant difference.

Results

Patient's characteristics

The baseline clinicopathological characteristics of all 2268 CRC patients from four centres are comprehensively presented in Table 1 and Table S1, Supplemental Digital Content 1, http://links.lww.com/JS9/B923. Among them, 1258 (55.5%) were male, and the median age (interquartile range) was 62.0 (52.0-69.0) years. The majority of patients ($n = 1846$, 81.4%) were diagnosed with stage II or III CRC.

Table 1

Clinicopathological characteristics of patients with CRC in the training and validation cohorts.

Characteristics	Training cohort ($n = 956$)	Internal validation cohort ($n = 240$)	External validation cohort 1 ($n = 509$)	External validation cohort 2 ($n = 203$)	External validation cohort 3 ($n = 360$)
Age (year), median (IQR)	61 (52–68)	61 (51–68)	64 (56–71)	60 (52–69)	60 (52–68)
Sex, *N* (%)					
Female	444 (46.4)	110 (45.8)	208 (40.9)	92 (45.3)	156 (43.3)
Male	512 (53.6)	130 (54.2)	301 (59.1)	111 (54.7)	204 (56.7)
Locations, *N* (%)					
Right colon	332 (34.7)	85 (35.4)	141 (27.7)	52 (25.6)	203 (56.4)
Left colon	283 (29.6)	70 (29.2)	143 (28.1)	60 (29.6)	156 (43.3)
Rectum	341 (35.7)	85 (35.4)	225 (44.2)	91 (44.8)	1 (0.3)
CEA level, *N* (%)					
≤5 (normal)	598 (62.6)	149 (62.1)	332 (65.2)	145 (71.4)	215 (59.7)
>5 (abnormal)	358 (37.4)	91 (37.9)	177 (34.8)	58 (28.6)	145 (40.3)
Differentiation, *N* (%)					
High	7 (0.7)	1 (0.4)	4 (0.8)	69 (34.0)	11 (3.1)
Middle	776 (81.2)	195 (81.3)	425 (83.5)	109 (53.7)	246 (67.8)
Low	173 (18.1)	44 (18.3)	80 (15.7)	25 (12.3)	103 (26.7)
pT stage, *N* (%)					
T1	16 (1.7)	5 (2.1)	16 (3.1)	8 (3.9)	17 (4.7)
T2	125 (13.1)	32 (13.3)	75 (14.7)	38 (18.7)	36 (10.0)
T3	273 (28.6)	66 (27.5)	371 (72.9)	140 (69.0)	245 (68.1)
T4	542 (56.7)	137 (57.1)	47 (9.2)	17 (8.4)	62 (17.2)
pN stage, *N* (%)					
N0	534 (55.9)	122 (50.8)	274 (53.8)	121 (59.6)	212 (58.9)
N1	246 (25.7)	61 (25.4)	147 (28.9)	67 (33.0)	93 (25.8)
N2	176 (18.4)	57 (23.8)	88 (17.3)	15 (7.4)	55 (15.3)
Stage, *N* (%)					
I	120 (12.6)	27 (11.3)	72 (14.1)	41 (20.2)	43 (11.9)
II	404 (42.3)	93 (38.8)	201 (39.5)	80 (39.4)	154 (42.8)
III	380 (39.7)	105 (43.8)	228 (44.8)	82 (40.4)	119 (33.1)
IV	52 (5.4)	15 (6.3)	8 (1.6)	0	44 (12.2)
LVI, *N* (%)					
Absent	757 (79.2)	182 (75.8)	390 (76.6)	187 (92.1)	284 (78.6)
Present	199 (20.8)	58 (24.2)	119 (23.4)	16 (7.9)	76 (78.6)
PNI, *N* (%)					
Absent	852 (89.1)	211 (87.9)	338 (66.4)	191 (94.1)	280 (78.6)
Present	104 (10.9)	29 (12.1)	171 (33.6)	12 (5.9)	80 (78.6)

χ^2 or Fisher's exact tests, were used to compare the differences in categorical variables, whereas student *t*-test or Mann–Whitney U test was used to compare the differences in continuous variables, as appropriate.

CEA, carcinoembryonic antigen; CRC, colorectal cancer; IQR, interquartile range; LVI, lymphovascular invasion; PNI, perineural invasion.

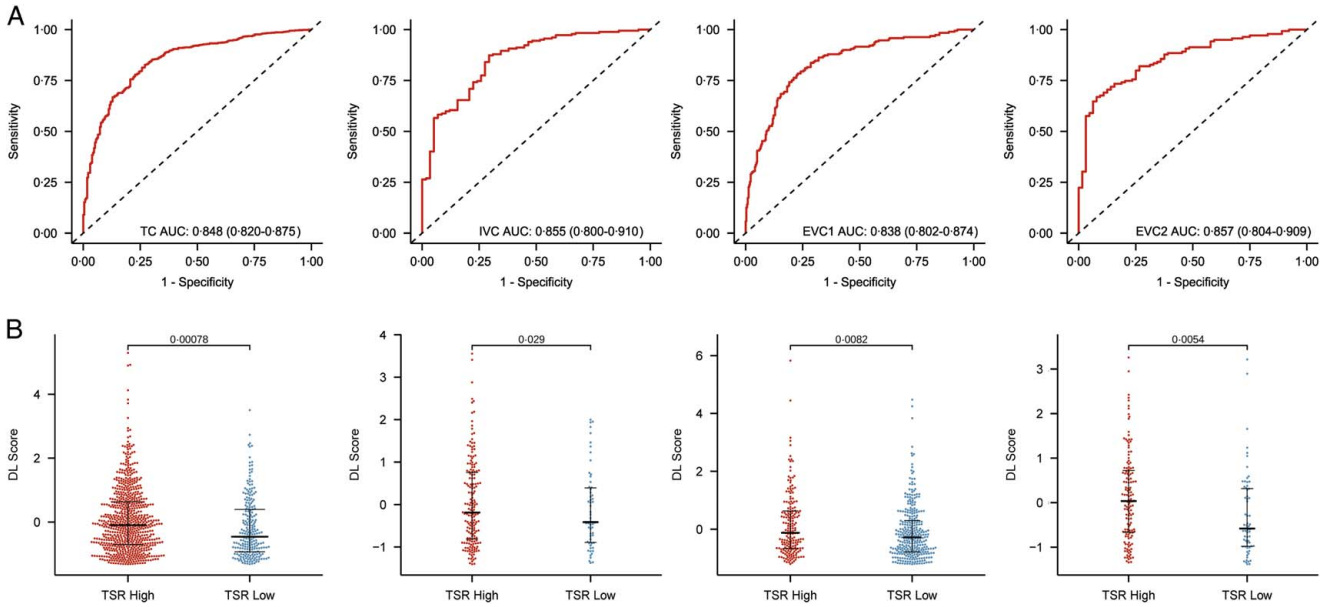

Figure 2. Diagnostic accuracy of the multitask deep learning model in the training and validation cohorts. (A) The receiver operating characteristic (ROC) curves of multitask deep learning model for predicting TSR status in each cohort; (B) Multitask deep learning score of high and low-TSR status in each cohort. AUC, area under the receiver operating characteristic curve; DL, Deep learning; EVC1, external validation cohort 1; EVC2, external validation cohort 2; IVC, internal validation cohort; TC, training cohort; TSR, tumour-stroma ratio.

Development and validation of the deep learning model for TSR prediction

To simultaneously predict TSR status and OS, we trained an MDL model based on preoperative portal-venous CT images. Figure S3, Supplemental Digital Content 1, http://links.lww.com/JS9/B923 illustrates two representative cases, displaying CT images alongside visualization of network-predicted TSR and OS visualizations.

The ability of MDL model for classifying high versus low TSR was shown to have an AUC of 0.848 (95% CI, 0.820–0.875) in the TC (Fig. 2A). This model demonstrated consistent discrimination for predicting TSR status in the IVC, EVC1 and EVC2, with AUCs of 0.855 (95% CI, 0.800–0.910), 0.838 (95% CI, 0.802–0.874), and 0.857 (95% CI, 0.804–0.909), respectively (Fig. 2A and Table S3, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). We further confirmed that the MDL score was significantly higher in the high-TSR group that those in the low-TSR group within each cohort (all $P < 0.001$) (Fig. 2B).

Prognostic value of the multitask deep learning model

We first confirmed that the TSR status, as determined by histopathological assessment, significantly correlated with both OS and DFS in the training and three validation cohorts (all $P < 0.05$)

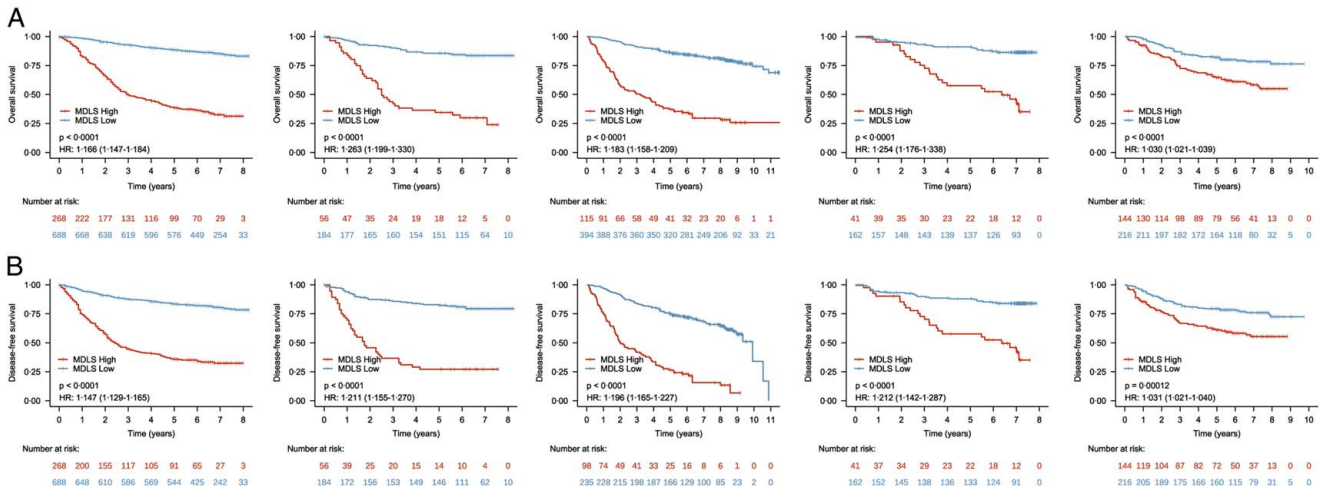

Figure 3. Kaplan–Meier analyses of overall survival (A) and disease-free survival (B) according to dichotomized multitask deep learning score (MDLS) in patients with colorectal cancer.

Table 2

Multivariable cox regression analysis of overall survival and disease-free survival in patients with colorectal cancer.

	Overall survival		Disease-free survival	
Variables	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Training cohort				
DLS[a]	1.154 (1.132–1.175)	< 0.0001	1.132 (1.111–1.153)	< 0.0001
Age (≥ 60 vs. <60 years)	2.088 (1.609–2.708)	< 0.0001	1.639 (1.283–2.094)	< 0.0001
CEA (elevated vs. normal)	1.353 (1.061–1.725)	0.015	1.342 (1.069–1.686)	0.011
N stage (N + vs. N0)	2.715 (2.014–3.661)	< 0.0001	2.441 (1.848–3.223)	< 0.0001
Stage (III + IV vs. I + II)	3.205 (2.191–4.689)	< 0.0001	2.716 (1.876–3.93)	< 0.0001
LVI (positive vs. negative)	1.553 (1.162–2.076)	0.003	1.524 (1.159–2.005)	0.002
Internal validation cohort				
DLS[a]	1.264 (1.195–1.338)	< 0.0001	1.194 (1.134–1.257)	< 0.0001
CEA (elevated vs. normal)	2.606 (1.569–4.329)	0.0002	1.821 (1.127–2.942)	0.014
N stage (N + vs. N0)	2.069 (1.144–3.743)	0.016	2.125 (1.198–3.768)	0.009
Stage (III + IV vs. I + II)	3.098 (1.572–6.104)	0.001	2.554 (1.298–5.025)	0.006
LVI (positive vs. negative)	2.072 (1.192–3.602)	0.009	—	—
External validation cohort 1				
DLS[a]	1.183 (1.15–1.216)	< 0.0001	1.171 (1.139–1.204)	< 0.0001
Age (≥ 60 vs. <60 years)	1.554 (1.078–2.24)	0.018	1.629 (1.132–2.345)	0.009
CEA (elevated vs. normal)	1.419 (1.002–2.009)	0.049	1.460 (1.037–2.055)	0.030
External validation cohort 2				
DLS[a]	1.209 (1.132–1.293)	< 0.0001	1.160 (1.093–1.232)	< 0.0001
Age (≥ 60 vs. <60 years)	3.323 (1.560–7.080)	0.002	3.074 (1.520–6.214)	0.002
CEA (elevated vs. normal)	1.944 (1.039–3.636)	0.037	1.943 (1.068–3.536)	0.029
LVI (positive vs. negative)	2.897 (1.333–6.295)	0.007	2.869 (1.375–5.987)	0.005
External validation cohort 3				
DLS[a]	1.013 (1.000–1.025)	0.044	1.011 (1.015–1.223)	0.048
CEA (elevated vs. normal)	1.536 (1.022–2.309)	0.039	1.523 (1.030–2.251)	0.035
Stage (III + IV vs. I + II)	4.174 (1.894–9.200)	0.0004	3.968 (1.882–8.367)	0.0003

[a]Continuous variable.

CEA, carcinoembryonic antigen; DLS, deep learning-based imaging signature; HR, hazard ratio; LVI, lymphovascular invasion.

(Figure S4, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). Subsequently, we evaluated the prognostic capacity of the MDL model. Remarkably, the model exhibited C-index values of 0.775 (95% CI, 0.745–0.804), 0.758 (95% CI, 0.693–0.823), 0.779 (95% CI, 0.739–0.819), and 0.757 (95% CI, 0.739–0.819) in the TC, IVC, EVC1 and EVC2, respectively.

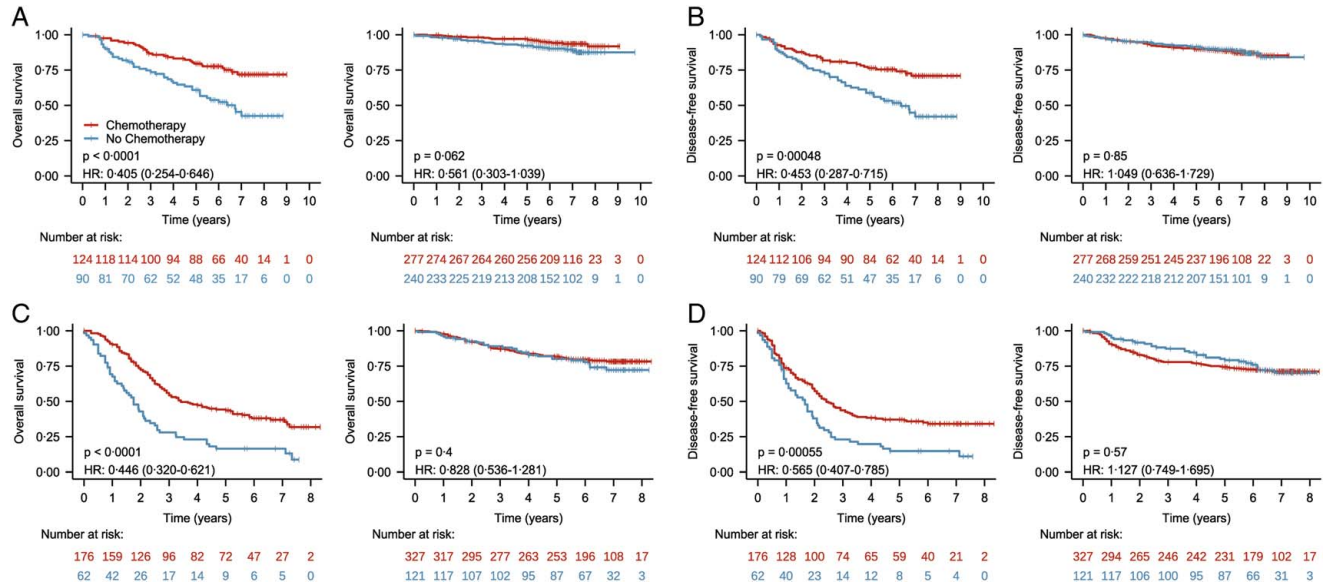

Figure 4. Relationship between the multitask deep learning score (MDLS) and overall survival (OS) and disease-free survival (DFS) in stage II (A for OS, B for DFS) and III (C for OS, D for DFS) patients who received or did not receive adjuvant chemotherapy. HR, hazard ratio.

By contrast, the model showed inferior performance for OS prediction in EVC3, with the C-index of 0.672 (0.622–0.721), where TSR data were unavailable. Kaplan–Meier survival curves underscored significant associations between MDL scores (MDLS) and distinct survival outcomes within each cohort(all $P < 0.001$)(Fig. 3 and Table S4, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). Similar trends were obtained for the DFS in each cohort (Fig. 3 and Table S4, Supplemental Digital Content 1, http://links.lww.com/JS9/B923).

We performed multivariate Cox regression analysis, with adjustments for clinicopathological variables, and found that the MDLS imaging biomarker remained a robust and independent prognostic factor for both OS and DFS across all cohorts (Table S5-6, Supplemental Digital Content 1, http://links.lww.com/JS9/B923 and Table 2). Stratified analyses further demonstrated that the performance of MDL model was not affected by diverse clinicopathological variables (Figure S5-6, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). Similar results were observed for IBS (Table S7, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). Moreover, an integrated model combining MDL model and clinicopathological features, including age, CEA level, N stage, LVI status, was built for individualized prediction of OS and DFS in the TC (Figure S7-8, Supplemental Digital Content 1, http://links.lww.com/JS9/B923 and Table 2). Notably, the nomogram exhibited significantly enhanced prognostic accuracy across all cohorts (Table S8, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). Correspondingly, prediction error curves consistently depicted lower prediction errors for the integrated nomogram in contrast to the DL model and clinicopathological model (Figure S9, Supplemental Digital Content 1, http://links.lww.com/JS9/B923).

Additionally, the NRI analysis revealed that the integration of MDLS into the nomogram performed satisfactorily in all cohorts, indicating improved classification accuracy for survival prediction (Table S9, Supplemental Digital Content 1, http://links.lww.com/JS9/B923). Calibration curves at 1-year, 3-year, or 5-year intervals exhibited favourable agreement between model estimations and actual observations across all cohorts (Figure S10, Supplemental Digital Content 1, http://links.lww.com/JS9/B923).

Predictive value of MDLS for chemotherapy response

To explore the predictive significance of the MDLS, we examined the correlation between MDLS and survival outcomes among stage II and III patients who either receive or did not receive adjuvant chemotherapy. Among the total 1846 patients with stage II and III CRC, 1417 (76.8%) have available post-surgery treatment information.

We found that for patients in the high-MDLS group, adjuvant chemotherapy was associated with an improved survival in both stage II [for OS, hazard ratio (HR) 0.405 (95% CI, 0.254–0.646), $P < 0.0001$] and stage III disease [HR0.446 (95% CI 0.320–0.621), $P < 0.001$](Fig. 4). Conversely, for patients in the low-MDLS groups, adjuvant chemotherapy did not affect survival in either stage II [for OS, HR0.561 (95% CI, 0.303–1.039), $P = 0.062$] or stage III disease [HR0.828 (95% CI, 0.536–1.281), $P = 0.40$]. Accordingly, an interaction test was conducted between the MDLS-defined risk groups and chemotherapy, confirming a significant interaction regarding the impact on OS and DFS in both stage II and III disease (all P values < 0.001). All these results suggest that the MDLS has predictive potential for chemotherapy benefits.

Superiority verification of proposed method

As show in Table S10, Supplemental Digital Content 1, http://links.lww.com/JS9/B923, multitask learning significantly improved TSR prediction in the training and three validation cohorts, which substantially increased AUCs to 0.838–0.857 from 0.799–0.827 with single-task learning. Additionally, multitask learning achieved better OS prediction than single-task learning, with C-index increased from 0.680–0.705 to 0.757–0.779 across all four cohorts.

We also compared the proposed multiscale learning approach with single-scale learning for TSR status and survival prediction. In all four cohorts, multiscale learning significantly improved prediction of TSR status, which substantially increased AUCs to 0.838–0.857 from 0.700–0.761 with single-scale learning. Moreover, multiscale learning achieved better prediction of OS with C-indexes of 0.757–0.779 versus 0.602–0.641 for single-scale learning in three validation cohorts.

Furthermore, attributable to integrated advantages of the RFA and MMD-AE modules over multiscale features, the network architecture we proposed efficaciously enhances model performance, and surpassing that of currently prevalent networks models (AUCs for TSR status: from 0.650–0.809 to 0.838–0.857, and C-indexes for survival: from 0.564–0.695 to 0.757–0.779 in all validation cohorts).

Discussion

The tumour-stroma microenvironment is crucial to disease progression, and its composition can influence treatment response and outcomes. TSR, an assessment of tumour-stroma amount within the tumour based on HE-staining of surgical specimens, has emerged as a robust prognostic and predictive biomarker, particularly in CRC. In this retrospective multicohort study, we developed and validated a noninvasive CT image-based multitask deep learning model that enabled simultaneous prediction of TSR status and prognosis post radical surgery for CRC patients. We further confirmed that the prognostic value of the deep learning signature was independent of various clinicopathological factors. Intriguingly, the tumour-stroma imaging signature demonstrated the ability to predict response to adjuvant chemotherapy in stage II and III CRC, indicating its potential to aid treatment decisions and follow-up management for CRC patients.

The present histological evaluation of tumour stroma encounters limitations in tissue accessibility, spatial heterogeneity, and temporal dynamics. In contrast, radiological imaging, boasts the advantage of being noninvasive and can be repeatedly obtained during the treatment course. The association between radiomics features and tumour immune microenvironment has been extensively investigated[13–15]. For instance, Sun *et al.* correlated both intratumor and peritumor radiomics features with CD8 T cells expression, and found that the CD8 radiomics signature could predict clinical response and outcomes in immunotherapy-treated patients[15]. Similar radiomics approaches from tumour and its periphery have been employed to evaluate the immune cells within TIME, such as immunoscore[13], neutrophil-to-lymphocyte ratio (NLR)[14], and tumour-infiltrating lymphocytes[28], exhibiting moderate performance with AUCs ranging from 0.74 to 0.86. In terms of TSR, only a few studies have explored radiomics features as TSR prediction biomarkers[29,30]. However, the clinical relevance of these

findings was constrained by the relatively small sample size, limited extensive external validation, and lack of survival or chemotherapy benefit data. Furthermore, the requirement for domain knowledge in handcrafted feature engineering and the time-intensive nature of classic radiomics analysis manual labelling have hindered clinical application[17]. Here, we developed DL model to noninvasively predict and validate TSR in the TME across multiple centres with 2268 patients. In the validation cohorts, the model exhibited TSR prediction promising performance, with AUCs exceeding 0.83, indicating that DL holds potential to capture diverse information encompassing tumour spatial heterogeneity and TME aspects related to tumour chemosensitivity.

Beyond predicting TSR, our multitask DL model was found to be significantly associated with prognosis in CRC, achieving C-index higher than 0.75 for OS in the four cohorts with TSR data. The rationale for integrating the TSR and survival prediction tasks into a unifying model, rather than predicting outcomes directly[12], is that while distinct, these tasks are intricately interconnected. This connection is particularly evident in the context of CRC, where the HE-derived TSR status, as demonstrated in our study, exhibited a robust correlation with OS and DFS. Of note, the survival patterns for the MDL model were similar to those based on histological TSR status evaluation. The prognostic strength of the MDL model was independent of clinicopathological variables, indicating its potential to categorize patients into distinct risk groups beyond the current staging system. Additionally, the prognostic efficacy was improved by combining the MDL model with clinicopathological factors. All these results indicate that, by linking CT images with TSR status and survival, the MDL model might capture underlying biology behind the survival predictions, unaffected by various confounding factors for outcomes.

From a technical standpoint, we developed and trained a multiscale and multitask deep learning network, and found that this approach outperforms a traditional single-task or single-scale DL for predicting TSR and survival. This fact was further validated by the inferior C-index of the single-task DL model for predicting OS in the EVC3, where TSR data unavailable. Our primary motivation for employing the multitask learning approach stems from the substantial correlation between the TSR status and prognosis. This strategy enables the multitask model to grasp the interplay between these tasks by acquiring general feature representations and facilitating information sharing, thereby enhancing the cohesiveness and robustness of predictions. Additionally, the multitask model promotes information exchange and sharing among various tasks to reduce the amount of model parameters for each task, and mitigate the risk of overfitting. However, these multiscale and multi-mask DL models present additional complexity when compared to a simple convolutional neural network. This complexity arises due to the need for optimizing a significantly larger number of parameters, resulting in a high computational cost and slower convergence of the model. Moreover, the generalization capabilities of neural decision forests may be uncertain when trained on small datasets commonly found in medical applications. Therefore, it is crucial to address these challenges, appropriately adapt sophisticated deep learning techniques, and develop robust models that can reliably predict clinical outcomes in medical settings.

Adjuvant chemotherapy is considered a standard treatment for patients with stage III CRC, while controversial for stage II patients. Currently, the optimal selection criteria for suitable candidates for chemotherapy remains uncertain, and only a small subset of patients could benefit from adjuvant chemotherapy[31]. Given the demonstrated ability of the MDL model to risk-stratify patients in our study, there is a potential for the MDL signature to enhance adjuvant treatment decision-making. Specifically, patients with high-MDLS could significantly benefit from adjuvant chemotherapy, while low-MDLS patients gained no benefit. Employing this MDL model, low-risk patients could avoid unnecessary chemotherapy, while high-risk patients could receive chemotherapy or more aggressive regimens to improve outcomes. Prospective studies to assess the impact of MDLS-informed treatment decisions on patient outcomes are warranted, particularly when integrated with established clinicopathologic criteria and molecular biomarkers that may provide additional prognostic information.

Notably, the present study has several limitations. Firstly, inherent bias was inevitable due to the retrospective nature and distribution differences, although multiple external validations were performed to improve reliability. Secondly, utilizing only the largest slice from pretreatment venous phase CT may not represent the entire tumour, introducing potential analysis bias. Thirdly, the use of adjuvant chemotherapy was not randomized, making it susceptible to selection biases. Finally, the biological significance of DL features warrants further investigation. Ultimately, prospective randomized clinical trials are necessary to validate the generalizability and clinical utility of our DL model.

Conclusion

In conclusion, we successfully developed and validated a multitask DL model utilizing preoperative CT images, that allows noninvasive evaluation of tumour-stroma microenvironment, particularly the TSR, as well as the clinical outcomes in patients with CRC. Moreover, the proposed DL model could be used to identify individuals who might benefit from adjuvant chemotherapy in stage II and III CRC. Further prospective randomized trials will be warranted to confirm its clinical applicability in refining prognosis and inform treatment decision in patients with CRC.

Ethical approval

Ethical approval for this study (Approval No., KY-N- 2022-004-01) was provided by the Ethical Committee of Guangdong Provincial People's Hospital, China on 21 January 2022.

Consent

Informed consent was waived owing to the retrospective nature of the study.

Source of funding

This study was supported by the Key-Area Research and Development Program of Guangdong Province, China (No.2021B0101420006), Regional Innovation and Development Joint Fund of National Natural Science Foundation of China (No. U22A20345), Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (No. 2022B1212010011), National Science Fund for Distinguished Young Scholars of China (No. 81925023),

National Natural Science Foundation of China (No. 82001789, 82102019, 82171923, 82202267, 82271946, 82371952), High-level Hospital Construction Project (No. DFJHBF202105), and the Applied Basic Research Projects of Shanxi Province, China, Outstanding Youth Foundation (No. 202103021222014).

Author contribution

Y.C., K.Z., X.Y., T.T., L.W., and Z.L. conceived and designed this study. Y.C., K.Z., M.X., and Y.M. collected the clinical dataset and conducted annotation. Y.C., K.Z., C.H., W.S., and L.W. performed data preprocessing and built the deep learning models. Y.C., K.Z., X.L.W. processed and analyzed the data. Y.C., X.M., X.Y., T.T., and L.W. prepared the initial manuscript draft, and other authors revised the manuscript. Z.L. critically revised the manuscript and provided consultation. All authors contributed to the writing of the manuscript, had full access to all the data in the study, approved the final manuscript, and had the final responsibility for the decision to submit for publication.

Conflicts of interest disclosure

All authors declare no competing interests.

Research registration unique identifying number (UIN)

This study was also approved by Chinese Clinical Trial (ChiCTR20000635734) https://www.chictr.org.cn/showproj.html?proj=184888.

Guarantor

Xiaotang Yang, Tong Tong, Lei Wu, Zaiyi Liu.

Data transparency statement

The datasets that support the findings of this study are available through data access agreement from the corresponding authors. De-identified clinical data will be provided on reasonable request. The image data are not publicly available because they contain sensitive information that could compromise patient privacy. Associated codes to process and analyze data are available on GitHub https://github.com/WuLei-MedIA/MDL4TSR_Survival].

Provenance and peer review

Not commissioned, externally peer-reviewed.

References

[1] Sung H, Ferlay J, Siegel RL, *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209–49.
[2] Weiser MR, Hsu M, Bauer PS, *et al*. Clinical calculator based on molecular and clinicopathologic characteristics predicts recurrence following resection of stage I-III colon cancer. J Clin Oncol 2021;39:911–9.
[3] Punt CJ, Koopman M, Vermeulen L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. Nat Rev Clin Oncol 2017;14:235–46.
[4] Kobayashi H, Enomoto A, Woods SL, *et al*. Cancer-associated fibroblasts in gastrointestinal cancer. Nat Rev Gastroenterol Hepatol 2019;16:282–95.
[5] Nicolas AM, Pesic M, Engel E, *et al*. Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer. Cancer Cell 2022;40:168–184 e113.
[6] van Wyk HC, Roseweir A, Alexander P, *et al*. The relationship between tumor budding, tumor microenvironment, and survival in patients with primary operable colorectal cancer. Ann Surg Oncol 2019;26:4397–404.
[7] van Wyk HC, Park JH, Edwards J, *et al*. The relationship between tumour budding, the tumour microenvironment and survival in patients with primary operable colorectal cancer. Br J Cancer 2016;115:156–63.
[8] Yuan Y. Spatial Heterogeneity in the Tumor Microenvironment. Cold Spring Harb Perspect Med 2016;6:a026583.
[9] Courrech Staal EF, Smit VT, van Velthuysen ML, *et al*. Reproducibility and validation of tumour stroma ratio scoring on oesophageal adenocarcinoma biopsies. Eur J Cancer 2011;47:375–82.
[10] Lambin P, Leijenaar RTH, Deist TM, *et al*. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.
[11] Huang YQ, Liang CH, He L, *et al*. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 2016;34:2157–64.
[12] Dai W, Mo S, Han L, *et al*. Prognostic and predictive value of radiomics signatures in stage I-III colon cancer. Clin Transl Med 2020;10:288–93.
[13] Jiang Y, Wang H, Wu J, *et al*. Noninvasive imaging evaluation of tumor immune microenvironment to predict outcomes in gastric cancer. Ann Oncol 2020;31:760–8.
[14] Huang W, Jiang Y, Xiong W, *et al*. Noninvasive imaging of the tumor immune microenvironment correlates with response to immunotherapy in gastric cancer. Nat Commun 2022;13:5095.
[15] Sun R, Limkin EJ, Vakalopoulou M, *et al*. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol 2018;19:1180–91.
[16] Berenguer R, Pastor-Juan MDR, Canales-Vazquez J, *et al*. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 2018;288:407–15.
[17] Li X, Zhang S, Zhang Q, *et al*. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. Lancet Oncol 2019;20:193–201.
[18] Chilamkurthy S, Ghosh R, Tanamala S, *et al*. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018;392:2388–96.
[19] Jiang X, Zhao H, Saldanha OL, *et al*. An MRI deep learning model predicts outcome in rectal cancer. Radiology 2023;307:e222223.
[20] Skrede OJ, De Raedt S, Kleppe A, *et al*. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet 2020;395:350–60.
[21] Jiang Y, Liang X, Han Z, *et al*. Radiographical assessment of tumour stroma and treatment outcomes using deep learning: a retrospective, multicohort study. Lancet Digit Health 2021;3:e371–82.
[22] Jiang Y, Zhang Z, Wang W, *et al*. Biology-guided deep learning predicts prognosis and cancer immunotherapy response. Nat Commun 2023;14:5135.
[23] Jiang Y, Zhang Z, Yuan Q, *et al*. Predicting peritoneal recurrence and disease-free survival from CT images in gastric cancer with multitask deep learning: a retrospective study. Lancet Digit Health 2022;4:e340–50.
[24] Amin MB, Greene FL, Edge SB, *et al*. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 2017;67:93–9.
[25] Mathew G, Agha R. for the STROCSS Group.. STROCSS 2021: Strengthening the Reporting of cohort, cross-sectional and case-control studies in Surgery. Int J Surg 2021;96:106165.
[26] Zhao K, Li Z, Yao S, *et al*. Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer. EBioMedicine 2020;61:103054.
[27] van Pelt GW, Sandberg TP, Morreau H, *et al*. The tumour-stroma ratio in colon cancer: the biological role and its prognostic impact. Histopathology 2018;73:197–206.

[28] Su GH, Xiao Y, Jiang L, *et al*. Radiomics features for assessing tumor-infiltrating lymphocytes correlate with molecular traits of triple-negative breast cancer. J Transl Med 2022;20:471.
[29] Meng Y, Zhang H, Li Q, *et al*. Magnetic Resonance Radiomics and Machine-learning Models: An Approach for Evaluating Tumor-stroma Ratio in Patients with Pancreatic Ductal Adenocarcinoma. Acad Radiol 2022;29:523–35.
[30] Cai C, Hu T, Gong J, *et al*. Multiparametric MRI-based radiomics signature for preoperative estimation of tumor-stroma ratio in rectal cancer. Eur Radiol 2021;31:3326–35.
[31] Grant RRC, Khan TM, Gregory SN, *et al*. Adjuvant chemotherapy is associated with improved overall survival in select patients with Stage II colon cancer: A National Cancer Database analysis. J Surg Oncol 2022; 126:748–56.