Case-control Study

INTERNATIONAL JOURNAL OF SURGERY

OPEN

Development and validation of a deep learning model for predicting gastric cancer recurrence based on CT imaging: a multicenter study

Mengxuan Cao, MSc[a,f,g], Can Hu, MD[a,f,g], Feng Li, MD[b], Jingyang He, MSc[a,f,g], Enze Li, MSc[a,f,g], Ruolan Zhang, MSc[a,f,g], Wenyi Shi, MSc[c,d], Yanqiang Zhang, MD[a,f,g], Yu Zhang, MD[h], Qing Yang, MSc[a,f,g], Qianyu Zhao, MSc[a,f,g], Lei Shi, MD, PhD[c,e,*], Zhiyuan Xu, MD, PhD[a,f,g,*], Xiangdong Cheng, MD, PhD[a,f,g,*]

Introduction: The postoperative recurrence of gastric cancer (GC) has a significant impact on the overall prognosis of patients. Therefore, accurately predicting the postoperative recurrence of GC is crucial.

Methods: This retrospective study gathered data from 2813 GC patients who underwent radical surgery between 2011 and 2017 at two medical centers. Follow-up was extended until May 2023, and cases were categorized as recurrent or nonrecurrent based on postoperative outcomes. Clinical pathological information and imaging data were collected for all patients. A new deep learning signature (DLS) was generated using pretreatment computed tomography images, based on a pretrained baseline (a customized Resnet50), for predicting postoperative recurrence. The deep learning fusion signature (DLFS) was created by combining the score of DLS with the weighted values of identified clinical features. The predictive performance of the model was evaluated based on discrimination, calibration, and clinical usefulness. Survival curves were plotted to investigate the differences between DLFS and prognosis.

Results: In this study, 2813 patients with GC were recruited and allocated into training, internal validation, and external validation cohorts. The DLFS was developed and assessed for its capability in predicting the risk of postoperative recurrence. The DLFS exhibited excellent performance with AUCs of 0.833 (95% CI: 0.809–0.858) in the training set, 0.831 (95% CI: 0.792–0.871) in the internal validation set, and 0.859 (95% CI: 0.806–0.912) in the external validation set, along with satisfactory calibration across all cohorts ($P > 0.05$). Furthermore, the DLFS model significantly outperformed both the clinical model and DLS ($P < 0.05$). High-risk recurrent patients exhibit a significantly poorer prognosis compared to low-risk recurrent patients ($P < 0.05$).

Conclusions: The integrated model developed in this study, focusing on GC patients undergoing radical surgery, accurately identifies cases at high-risk of postoperative recurrence and highlights the potential of DLFS as a prognostic factor for GC patients.

Keywords: deep learning, gastric cancer, model, radiomics, recurrence

Introduction

Gastric cancer (GC) is a prevalent malignancy and ranks as the fifth most common cancer globally. Unfortunately, it is also the fourth leading cause of cancer-related deaths worldwide[1]. While there has been a noteworthy decrease in the incidence of stomach cancer in recent years, it remains the third leading cause of cancer-related mortality[2,3]. With advancements in medical technology and treatment paradigms, the current approach to treating GC is comprehensive, centering around surgical and chemotherapeutic

[a]Department of Gastric Surgery, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang, [b]School of Biomedical Engineering, ShanghaiTech University, Shanghai, [c]Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, [d]School of Molecular Medicine, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, [e]Department of Radiology, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, [f]Key Laboratory of Prevention, Diagnosis and Therapy of Upper Gastrointestinal Cancer of Zhejiang Province, Hangzhou, [g]Zhejiang Provincial Research Center for Upper Gastrointestinal Tract Cancer, Zhejiang Cancer Hospital, Hangzhou and [h]Zhejiang Hospital of Traditional Chinese Medicine, Hangzhou, Zhejiang, People's Republic of China

Mengxuan Cao, Can Hu, and Feng Li contributed equally to this work.

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

*Corresponding author. Address: Department of Gastric Surgery, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, People's Republic of China. Tel.: +8613968032995. E-mail: chengxd@zjcc.org.cn (X. Cheng), and Tel.: +8613757198610. E-mail: xuzy@zjcc.org.cn (Z. Xu); Department of Radiology, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, People's Republic of China Tel.: 15988872208. E-mail: shilei@zjcc.org.cn (L. Shi).

International Journal of Surgery (2024) 110:7598–7606

Received 2 April 2024; Accepted 6 May 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 20 June 2024

http://dx.doi.org/10.1097/JS9.0000000000001627

interventions[4,5]. Most patients can experience significant survival benefits with individualized treatments. However, for those who experience GC recurrence after surgery, the outcomes are often less than optimal. This results in significantly lower overall postoperative survival rates compared to patients without recurrence[6]. In China, over 60% of patients with advanced GC face recurrences and metastases following surgery[7].

Postoperative recurrence of gastric cancer (PRGC) is mostly the same as the initial occurrence, the clinical manifestations are relatively hidden, and patients often have no conscious symptoms[8]. Clinically, the recurrence probability is mainly estimated by pTNM stage and regular follow-up review to diagnose recurrence[9]. However, in the face of complex tumor microenvironment, many previous studies have questioned whether traditional TNM stage can still accurately predict patient prognosis and whether it can achieve individualized tumor treatment[10,11]. Nowadays, imaging examination is the main method of postoperative review for patients, including computed tomography (CT), endoscopic ultrasonography (EUS), MRI, and positron emission tomography (PET-CT)[5,12]. With continuous innovation in medical technology and advances in medical devices, imaging tests can detect more subtle evidence of relapse than ever before. Therefore, it is necessary to develop a reliable predictive tool to predict the recurrence and metastasis of GC after surgery.

Recently, radiomics and deep learning have emerged as an emerging class of tumor analysis and prediction methods, showing great promise in oncology practice[13–15]. A retrospective study has shown that the combination of radiomics, histological type, and tumor size is an effective method for predicting lymph node metastasis in GC[16]. Another multicenter study suggests utilizing deep learning to predict resistance of locally advanced gastric cancer (LAGC) to neoadjuvant chemotherapy[17]. However, to the best of our knowledge, there is currently a lack of large multicenter studies using deep learning to predict PRGC based on radiomics in our country.

In this study, we aim to integrate radiomics with clinical factors and utilize a deep learning approach to establish a fused prediction model for postoperative recurrence after radical gastrectomy.

HIGHLIGHTS

- The predictive performance of fusion model is significantly better than that of radiomics model and clinical model.
- The survival time of patients with high-risk recurrence of gastric cancer identified by fusion prediction model was significantly lower than that of patients with low-risk recurrence.
- Adjuvant chemotherapy had no effect on the predictive performance of the model, and the predictive performance was higher.

Methods

Materials

We retrospectively analyzed 2813 GC patients from two large medical centers who underwent radical gastrectomy at a medical center from January 2011 to January 2017. Among them, 2481 patients were sourced from a cancer treatment center, with 1774 patients randomly assigned to the training group and 707 patients to the internal validation group (ratio 7:3). Additionally, 332 patients from another medical center were included as an external validation set. The inclusion criteria of the patients in this study included: (1) primary GC confirmed by postoperative pathological examination; (2) radical resection after R0 resection of GC with negative upper and lower margins; (3) lymph node dissection with D2 and above; (4) no recurrence or death within 2 months after radical surgery; (5) CT with abdominal contrast was performed 1 week before surgery; (6) complete CT image data and clinical treatment records and follow-up data (regular outpatient follow-up and telephone follow-up) (Supplementary Figure S1, Supplemental Digital Content 1, http://links.lww.com/JS9/C798).

Baseline clinicopathological data were collected, including age, sex, nerve invasion, vascular tumor thrombus, pTNM stage, relapse status, tumor length, differentiation, and levels of pretreatment tumor markers (CEA, AFP, CA19-9, CA125, and CA72-4). According to the AJCC/UICC 8th Edition TNM staging criteria, the TNM stage was confirmed via pathological evaluation. The primary adjuvant chemotherapy regimens included SOX, XELOX, FOLFOX, FLOT, and others. The study was approved by the ethics committees of the participating center (IRB-2024-220), and the study was conducted in accordance with the principles of the Declaration of Helsinki. The work has been reported in line with the strengthening the reporting of cohort, cross-sectional, and case–control studies in surgery (STROCSS) criteria[18] (Supplemental Digital Content 2, http://links.lww.com/JS9/C799).

Tumor segmentation

All patients underwent enhanced CT examination within 1 week before surgery. Tumor segmentation was performed by three experienced radiologists using ITK-SNAP software (version 3.8, http://www.itksnap.org). Since GC can be clearly distinguished from surrounding normal gastric tissue in enhanced portal phase CT images, all mapping is performed along tumor boundaries in portal phase CT images. In addition, taking into account the holistic nature of the tumor, this study mapped tumor lesions at all levels on CT. If an uncertain tumor lesion is encountered, the three physicians will focus on the discussion about the division of the ROI to ensure the accuracy of the delineation.

Data preparation

3D processed VOIs of each patient are regarded as the model inputs. Prior to analysis, the inputs' pixel spacing is normalized to a uniform size of $1.0 \times 1.0 \times 1.0$ mm^3 and a filtering window of [−115 to 235] HU is then applied to them. Subsequently, the short side of 2D images on the z-axis is firstly zero padded to the length of long side and then resize them to 112×112. Furthermore, we normalized the pixel values of the images to the range of [−1 to 1]. This normalization process ensures consistency and facilitates optimal performance during the training phase. Additionally, to augment the training dataset, we employed three transformations (flipping, rotation, and Gaussian filters) on the images before feeding them into the network. By adopting these strategies, we aimed to enhance the robustness and reliability of our analysis pipeline.

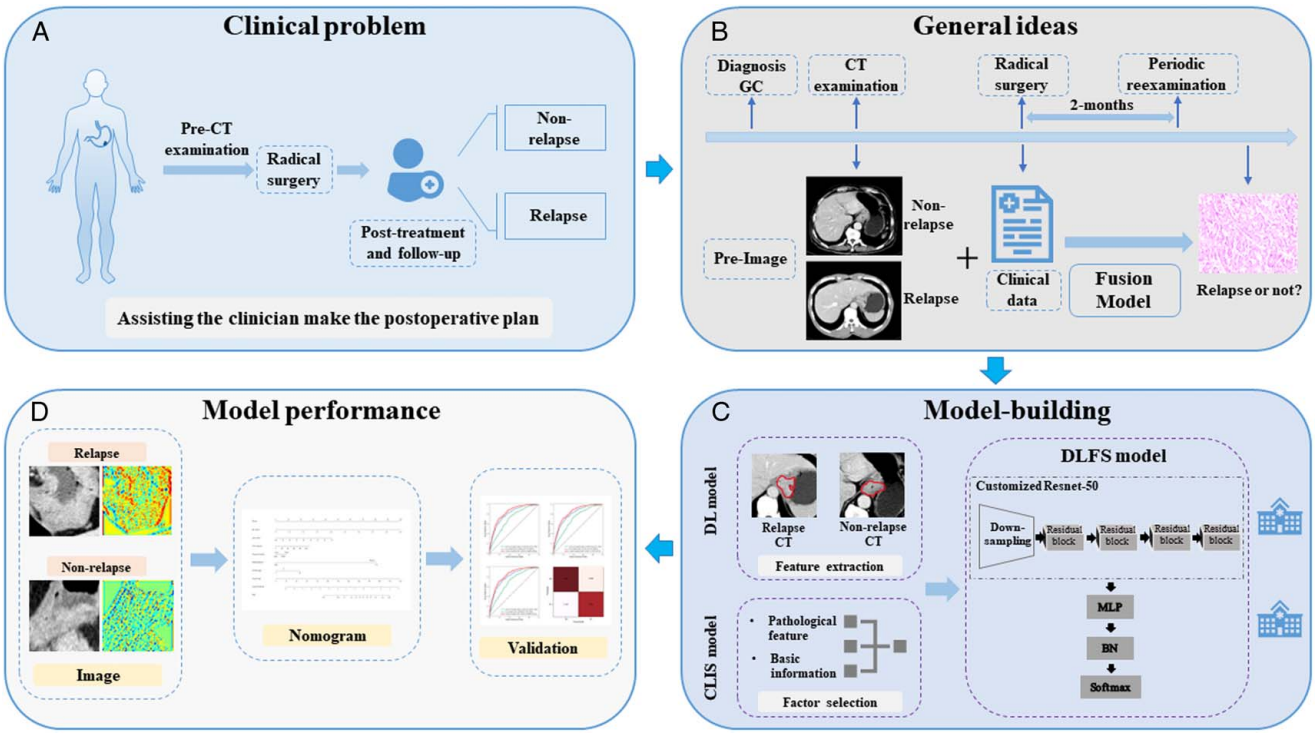

Figure 1. Research design concept and technical workflow diagram. The whole line of thought of the research, including the clinical problem (A), general ideas (B), model-building (C), model performance (D).

Construction of the deep learning fusion signature

Considering the importance of clinicopathologic characteristics, a clinical PRGC prediction model (Clinical Signature, CLIS) is constructed based on multivariate logistic regression (MLR) algorithm[19]. The most valuable features were identified from 17 relevant clinicopathologic factors through the application of the LASSO (Least Absolute Shrinkage and Selection Operator) method on the training cohort.

Our entire deep learning network for PRGC prediction (DL Signature, DLS) primarily consists of two components: a pre-trained customized Resnet50[20] designed for tumor regression grade classification on LAGC patients and a newly added multilayer perceptron (MLP) sub-network before the softmax layer. In order to expedite model convergence and balance its performance on imbalanced class datasets, we have frozen all weight parameters of the previously customized Resnet50 model, which has similar classification tasks and datasets, and solely train the extra added MLPs. For model interpretation, the feature heatmaps of our network are further utilized during the inference stage on the two validation sets to visualize the most informative regions detected by the network, aiding in decision-making related to PRGC. Eventually, the fusion model [deep learning fusion signature (DLFS)] is built on the above calculated clinical features and DL score by MLR method ($P < 0.05$). A comprehensive workflow is employed as illustrated in Figure 1.

Statistical analysis

The DL network was implemented utilizing PyTorch (version 1.13.1). Nonparametric data with a non-normal distribution were analyzed using the Mann–Whitney U test. Continuous variables with a normal distribution were assessed using Student's t-test, while categorical variables were examined using the χ^2 test. All tests of statistical significance were two-sided and conducted at a significance level of $\alpha = 0.05$. To evaluate the performance of the models, receiver operating characteristic (ROC) analysis and precision-recall curves were employed. The continuous probability score, ranging from 0 to 1, was utilized as the basis for these analyses. Additionally, decision curve analysis (DCA) was conducted to assess the clinical utility of the prediction models. DCA quantifies the net benefit at various threshold probabilities, providing insights into the practical usefulness of the models. During the training stage, the model was developed with a batch size of 128. The initial learning rate was set to 0.0001. It was trained on two GeForce RTX 2080 Ti GPUs using the PyTorch framework. The training process spanned a maximum of 1000 epochs, with an early-stopping mechanism implemented to terminate training if no improvement was observed for 100 consecutive epochs.

Results

Statistical analysis of clinical characteristics

This study included a total of 2813 patients, comprising 2481 patients from a cancer treatment center [training cohort ($n = 1774$), internal validation cohort ($n = 707$)] and 332 patients from another medical center (external validation cohort). Table 1 presents detailed pathological characteristics of all patients. The overall median age (interquartile range) was 61.0 (54.0–68.0)

Table 1

Clinicopathologic characteristics of patients in the training and validation cohorts

	Training set ($N=1774$)			Internal validation set ($N=707$)			External validation set ($N=332$)		
Variables	**Relapse ($n=275$)**	**Nonrelapse ($n=1499$)**	***P***	**Relapse ($n=104$)**	**Nonrelapse ($n=603$)**	***P***	**Relapse ($n=36$)**	**Nonrelapse ($n=296$)**	***P***
Sex									
Male	200 (72.7)	1092 (72.8)	0.967	72 (69.2)	447 (74.1)	0.296	23 (63.9)	198 (66.9)	0.718
Female	75 (27.3)	407 (27.2)		32 (30.8)	156 (25.9)		13 (36.1)	98 (33.1)	
Vascular tumor thrombus									
Negative	111 (40.4)	839 (56.0)	**0.000**	41 (39.4)	314 (52.1)	**< 0.05**	10 (27.8)	140 (47.3)	**< 0.05**
Positive	164 (59.6)	660 (44.0)		63 (60.6)	289 (47.9)		26 (72.2)	156 (52.7)	
Nerve invasion									
Negative	94 (34.2)	775 (51.7)	**0.000**	34 (32.7)	267 (44.3)	**< 0.05**	14 (38.9)	184 (62.2)	**< 0.05**
Positive	181 (65.8)	724 (48.3)		70 (67.3)	336 (55.7)		22 (61.1)	112 (37.8)	
Tumor location									
Upper	74 (26.9)	362 (24.2)	**0.000**	22 (21.2)	164 (27.2)	**< 0.05**	11 (30.6)	54 (18.3)	0.055
Middle	82 (29.8)	315 (21.0)		30 (28.8)	84 (13.9)		8 (22.2)	109 (36.8)	
Lower	108 (39.3)	798 (53.2)		49 (47.1)	338 (56.1)		16 (44.4)	132 (44.6)	
Whole	11 (4.0)	24 (1.6)		3 (2.9)	17 (2.8)		1 (2.8)	1 (0.3)	
Differentiation									
Low	165 (60.0)	713 (47.6)	**0.000**	67 (64.4)	307 (50.9)	**< 0.05**	30 (83.3)	209 (70.6)	0.268
Medium	110 (40.0)	751 (50.1)		36 (34.6)	282 (46.8)		5 (13.9)	68 (23.0)	
High	0 (0.0)	35 (2.3)		1 (1.0)	14 (2.3)		1 (2.8)	19 (6.4)	
Preoperative chemotherapy									
Negative	240 (87.3)	1315 (87.7)	0.834	79 (76.0)	536 (88.9)	**0.000**	30 (83.3)	262 (88.5)	0.367
Positive	35 (12.7)	184 (12.3)		25 (24.0)	67 (11.1)		6 (16.7)	34 (11.5)	
Postoperative chemotherapy									
Negative	105 (38.2)	811 (54.1)	**0.000**	16 (15.4)	337 (55.9)	**0.000**	14 (38.9)	176 (59.5)	**< 0.05**
Positive	170 (61.8)	688 (45.9)		88 (84.6)	266 (44.1)		22 (61.1)	120 (40.5)	
pT stage									
T1/T2	31 (11.3)	478 (31.9)	**0.000**	20 (19.2)	207 (34.3)	**< 0.05**	8 (22.2)	135 (45.6)	**< 0.05**
T3/T4	244 (88.7)	1021 (68.1)		84 (80.8)	396 (65.7)		28 (77.8)	161 (54.4)	
pN stage									
N0	41 (14.9)	530 (35.4)	**0.000**	13 (12.5)	211 (35.0)	**0.000**	8 (22.2)	140 (47.3)	**< 0.05**
N1	34 (12.4)	269 (17.9)		13 (12.5)	90 (14.9)		5 (13.9)	47 (15.9)	
N2	65 (23.6)	318 (21.2)		25 (24.0)	124 (20.6)		5 (13.9)	41 (13.8)	
N3	135 (49.1)	382 (25.5)		53 (51.0)	178 (29.5)		18 (50.0)	68 (23.0)	
pTNM stage									
I	19 (6.9)	361 (24.1)	**0.000**	9 (8.7)	150 (24.9)	**0.000**	6 (16.7)	110 (37.2)	**< 0.05**
II	36 (13.1)	345 (23.0)		12 (11.5)	106 (17.6)		7 (19.4)	70 (23.6)	
III	220 (80.0)	793 (52.9)		83 (79.8)	347 (57.5)		23 (63.9)	116 (39.2)	
Age, median [IQR], y	59.00 [52.00–66.00]	62.00 [55.00–68.00]	**0.000**	57.00 [49.25–63.00]	60.00 [52.00–67.00]	**< 0.05**	63.50 [54.50–68.75]	64.00 [58.00–70.00]	0.658
Tumor length, median [IQR], cm	5.00 [3.80–7.00]	4.00 [2.50–6.00]	**0.000**	4.50 [3.35–5.86]	4.50 [3.00–6.10]	0.447	4.25 [3.00–6.75]	3.50 [2.20–5.50]	**< 0.05**
AFP, median [IQR], ng/ml	2.49 [1.74–3.69]	2.40 [1.68–3.45]	0.590	3.05 [2.28–4.21]	3.22 [2.46–4.56]	0.798	2.65 [2.01–3.44]	2.68 [2.00–3.64]	0.642
CEA, median [IQR], ng/ml	2.49 [1.35–4.71]	2.35 [1.40–4.30]	**< 0.05**	2.59 [1.52–4.46]	2.47 [1.54–3.86]	**< 0.05**	2.50 [1.45–4.03]	2.60 [1.53–4.38]	**< 0.05**
CA19-9, median [IQR], U/ml	11.66 [5.18–42.99]	10.78 [5.66–25.31]	**< 0.05**	13.36 [6.31–38.15]	8.34 [4.18–20.07]	0.883	4.56 [2.72–14.21]	7.68 [3.53–21.38]	0.651
CA125, median [IQR], U/ml	12.00 [8.78–19.20]	10.80 [7.70–15.70]	**< 0.05**	12.50 [8.60–17.15]	9.50 [6.80–14.75]	0.272	11.90 [7.40–13.80]	12.65 [6.70–16.10]	0.965
CA72-4, median [IQR], U/ml	2.97 [1.60–7.20]	2.13 [1.19–5.74]	0.874	2.58 [1.24–6.00]	2.24 [1.22–4.97]	0.221	2.12 [1.27–2.87]	2.12 [1.39–3.62]	0.253

Statistical significance ($P < 0.05$) values are in bold.

years, with 2032 patients (72.2%) being male. The majority of patients (76.7%) were diagnosed with stage II or III GC. Among all patients, 351 (12.5%) received preoperative neoadjuvant chemotherapy, and 1354 (48.1%) received postoperative adjuvant chemotherapy.

Selection of clinicopathological factors and establishment of clinical model

As shown in Table 1, this study included a total of 17 clinicopathological factors. In the training set, significant differences

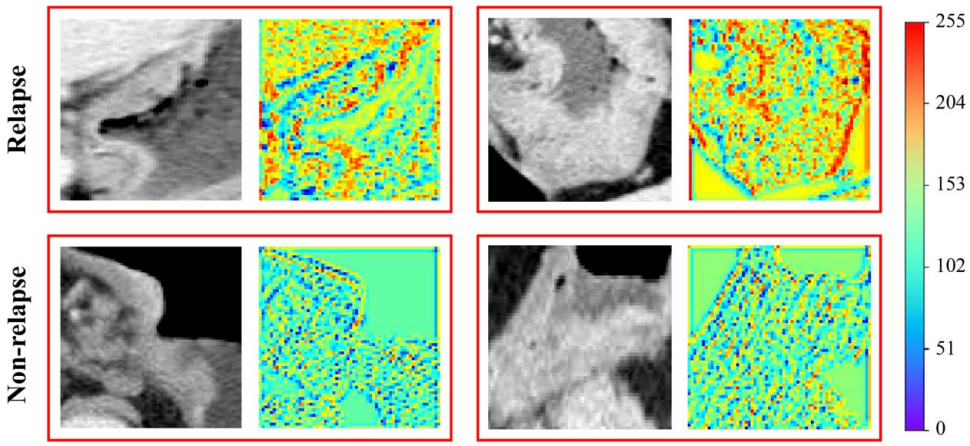

Figure 2. The visual output map. The tumor region is depicted through imaging features extracted via DLS, which are correlated with the intratumoral heterogeneity.

were observed between the recurrence group and nonrecurrence group for age, vascular tumor thrombus, nerve invasion, tumor length, postoperative adjuvant chemotherapy, tumor location, differentiation, CEA, CA19-9, CA125, pT stage, pN stage, and pTNM stage ($P < 0.05$). Subsequently, after adjusting for additional clinical factors and conducting LASSO regression analysis combined with radiomic features, the final CLIS included age, CEA, tumor location, differentiation, and pTNM stage.

Construction of models for predicting PRGC

The visual output map, depicting the tumor area, presents the extracted imaging characteristics through the DLS associated with intratumor heterogeneity. These heatmaps contain valuable information that can assist in predicting the PRGC. They demonstrate significant differences among different groups (Fig. 2). Their PRGC signatures (DLS, CLIS, and DLFS) were constructed based on CT-based DL extracted features, significant clinicopathologic characteristics, and the fused ones, respectively. This was achieved through internal fivefold cross-validation on the training set. Radiomic features based on deep learning combined with clinically independent predictors of PRGC. Finally, we developed a DLFS to predict recurrence that included DLS, age, CEA, tumor location, differentiation, and pTNM stage (Fig. 3A).

Validation of predictive performance of the models

Compared with the results of CLIS, DLS exhibited significantly better discriminatory ability than the clinical signature ($P < 0.0001$), with AUCs of 0.804 (95% CI: 0.777–0.831) and 0.704 (95% CI: 0.672–0.736) on the training set, with AUCs of 0.791 (95% CI: 0.747–0.835) and 0.679 (95% CI: 0.624–0.734) on the internal validation set, with AUCs of 0.808 (95% CI: 0.740–0.876) and 0.637 (95% CI: 0.530–0.744) on the external validation set, respectively (Fig. 3B, C and D).

Notably, the multimodality model, DLFS exhibited significantly improved ROCs compared to DLS (2.9, 4, and 5.1%, respectively) on the three datasets. In the training and internal validation set, DLFS showed an improvement in sensitivity of 9.5% and 16.4% compared to DLS, respectively. There was also an improvement of 10.2% and 10.6% compared to CLIS, respectively. However, the specificity decreased by 4.7% and 6% on those two datasets, respectively, compared to DLS. This decrease can be attributed to the poor performance of CLIS on those datasets (Supplementary Table S1, Supplemental Digital Content 1, http://links.lww.com/JS9/C798). Furthermore, the DLFS showed good consistency in the calibration curves for predicting PRGC in all three cohorts (Supplementary Figure S2, Supplemental Digital Content 1, http://links.lww.com/JS9/C798).

Additionally, the confusion matrices presented also showcase the performance of our predictive model across the three cohorts (Fig. 3E, F and G). The confusion matrix is a table used to evaluate the model's performance. In essence, a higher TP value indicates greater accuracy in the model. The model established in this study demonstrated satisfactory performance across all three cohorts, with high TP values for predicting patient recurrence at 0.844, 0.856, and 0.722. This suggests that the model possesses a notable capability to identify positive samples, indicating its effectiveness in recognizing instances of recurrence. In the external validation cohort, the presence or absence of preoperative neoadjuvant or postoperative adjuvant chemotherapy had no significant impact on the performance of the model. DLFS still demonstrated high precision in prediction, with the AUC values ranging from 0.829 to 0.889 (Supplementary Figure S3, Supplemental Digital Content 1, http://links.lww.com/JS9/C798).

Decision curve analysis

Supplement Figure S4 (Supplemental Digital Content 1, http://links.lww.com/JS9/C798) illustrates the decision curve analysis comparing the developed fusion model to the independently evaluated radiomics and clinical models. The results demonstrate that the fusion model exhibits superior net benefit compared to the radiomics and clinical models.

Survival analysis

We aimed to further investigate the association between DLFS features of this model and prognosis. Based on the DLFS model scores from the training cohort, we determined the optimal cutoff score for DLFS to be 0.60. Using this threshold, patients were categorized into two groups: DLFS high-risk (≥ 0.60) and

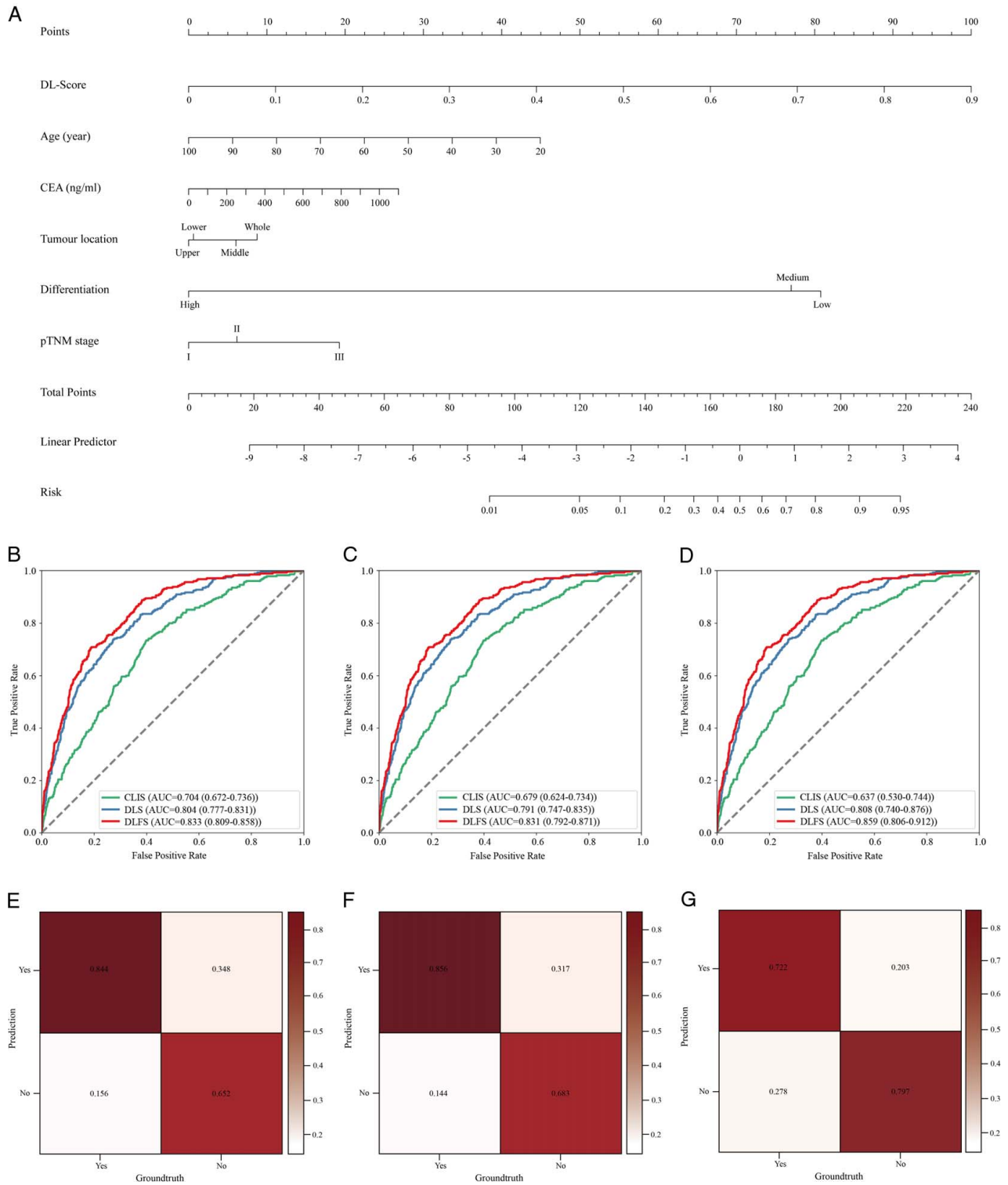

Figure 3. Construction and performance of deep learning fusion model (DLFS). A fusion signature of DLFS was built from the DL signature (DLS) and clinical signature (CLIS), including age, tumor marker CEA, tumor location, differentiation and pTNM stage (A). Receiver operating characteristic (ROC) curves of the CLIS, DLS and DLFS for predicting postoperative recurrence of gastric cancer in training cohort (B), internal validation cohort (C), external validation cohort (D). The confusion matrices of the DLFS in the training cohort (E), internal validation cohort (F), and external validation cohort (G).

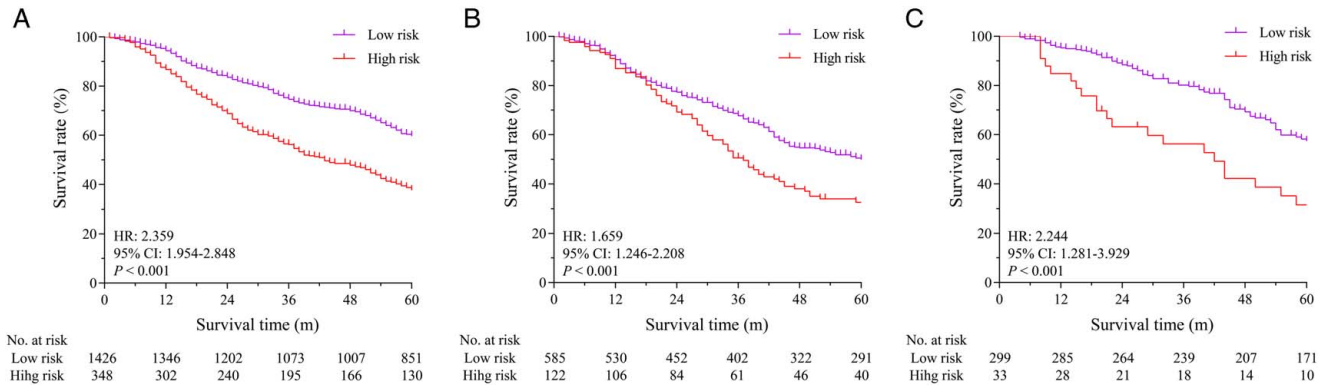

Figure 4. Kaplan–Meier analyses of overall survival. Survival difference between high-risk group (≥0.60) and low-risk group (<0.60) in the training cohort (A), internal validation cohort (B), and external validation cohort (C).

low-risk (<0.60) groups. Results revealed significant differences in prognosis between high-risk and low-risk recurrence patients in the training cohort, internal validation cohort, and external validation cohort ($P < 0.001$), with respective HR of 2.359 (95% CI: 1.954–2.848), 1.659 (95% CI: 1.246–2.208), and 2.244 (95% CI: 1.281–3.929) (Fig. 4).

Discussion

Postoperative recurrence of cancer is a complex process involving many factors. For recurrent patients, early identification and prevention are crucial for treatment[6]. Therefore, accurate prediction of postoperative recurrence in GC is of great importance in clinical practice. We collected data from two large medical centers and developed a deep-learning model based on radiomics to predict PRGC. This model performs exceptionally well in identifying high-risk patients for recurrence and was also significantly associated with prognosis. Furthermore, in terms of predictive performance, the fusion model outperforms models based solely on clinical or radiological features. This demonstrates our model as a personalized tool for predicting PRGC, guiding clinicians in their treatment decisions.

With the development of medical technology, many new technologies have been introduced into the clinic. In recent years, deep learning has become an indispensable concept in the field of artificial intelligence[21,22]. It automatically learns image features, eliminating the need for traditional manual feature extraction and making medical image analysis more automated and accurate[23]. The integration of radiomics and deep learning into clinical models has been proven to significantly improve the accuracy of predictions, especially in cancer diagnosis, treatment planning, and disease prognosis assessment[24–26]. A study developed a model that relies on preoperative CT images to predict prognosis and the benefits of adjuvant chemotherapy[20]. Cui *et al.*[27] established a deep learning radiomics nomogram (DLRN) to predict the response of LAGC patients to neoadjuvant chemotherapy (NACT). These studies provide essential references for the subsequent treatment of GC patients.

Currently, there is a lack of large-scale multicenter studies to predict PRGC in China. In their recent investigation, Jiang *et al.*[28] concentrated on the prediction of peritoneal recurrence in GCs. Employing deep learning features in conjunction with CT images, they aimed to forecast both peritoneal recurrence and survival, achieving promising results. In a previous study, a radiomics-based approach was developed to predict early recurrence in LAGC patients before surgery. However, their sample size was relatively small ($n = 669$), and the relationship with prognosis was not explored[29]. In our study, we have a larger sample size ($n = 2813$) and have demonstrated excellent predictive performance during validation, as well as established the association between DLFS and prognosis. Patients identified as high-risk by the DLFS, which we developed, exhibit a poorer prognosis compared to those with lower risk ($P < 0.05$), offering valuable insights for patient prognosis. As far as we know, our study is the first multicenter research in China that employs deep learning radiomics for predicting overall postoperative recurrence in GC on such a large-scale. Specially, the DLS for PRGC prediction was developed rely on our previous customized ResNet50 considering the similar feature space. The model was further fine-tuned efficiently with only extra added MLP block. Not only does it save training expenses, but it also further proves that the features extracted by our model have certain generalization capabilities. The utilization of provided DLS enables the extraction of imaging characteristics linked to intratumor heterogeneity, and this, in turn, contributes to the creation of a visualized output map showcasing the tumor area. This map holds significant promise in offering valuable insights for the prediction of PRGC.

In the CLIS, age, CEA, primary tumor location, differentiation, and pTNM stage were identified as important factors for predicting the risk of recurrence. Among them, we observed that tumor low differentiation is considered the most crucial factor for recurrence. Consistent with previous research, the lower the degree of differentiation, the higher the malignancy, exhibiting characteristics such as high invasiveness, rapid growth, and strong metastatic potential. Many prior prospective studies have demonstrated the impact of neoadjuvant and adjuvant chemotherapy on recurrence in advanced GC patients[30–32]. In the external validation cohort, we explored the impact of chemotherapy on DLFS prediction. The results show consistently high predictive accuracy, irrespective of chemotherapy administration. Contrary to previous reports, chemotherapy is not a key factor in this study. This may be attributed to the inclusion of a significant number of nonadvanced-stage patients (43.8%, $n = 1231$), who have a relatively lower recurrence rate (7.2%, $n = 89$), and rarely undergo adjuvant chemotherapy.

Subsequently, further investigation into treatment factors is warranted to provide clinical evidence.

Additionally, this study has some limitations. First, although the sample size is large and multicenter, inherent biases are unavoidable due to its retrospective nature and varying distribution. Therefore, further carefully designed prospective studies are needed to verify the clinical applicability of this model. Second, while we discussed the prognosis of recurrent patients, we did not establish related survival prediction models, and the results require further modeling and analysis. Furthermore, we did not find chemotherapy to be a crucial factor in preventing recurrence. We suspect that this may be linked to the study population and specific chemotherapy regimens, requiring more detailed subgroup analysis. Finally, our study was developed based on patients of Asian ethnicity. Since the epidemiology of GC in Eastern countries differs significantly from Western countries, further validation on patients from different ethnicities is necessary.

Conclusions

In summary, we developed and validated a CT-based model using deep learning and radiomics analysis for predicting postoperative recurrence in GC patients who underwent radical resection. The resulting DLFS integrates imaging features and clinical factors, exhibiting good performance in predicting treatment response and clinical outcomes. It offers valuable insights for tailoring treatment and follow-up strategies for GC patients. Nevertheless, future prospective studies are required to validate the clinical applicability of this DLFS.

Ethical approval

The study was approved by research ethics committee of the Zhejiang Cancer Hospital (IRB-2024- 220), and this study was undertaken in accordance with the World Medical Association-Declaration of Helsinki – ethical principles for medical research.

Consent

This study was approved by the ethics committee of each participating hospital. The requirement for informed consent was waived. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Source of funding

This study was supported by The National Key Research and Development Program of China (2021YFA0910100), Healthy Zhejiang One Million People Cohort (K-20230085), Postdoctoral Innovative Talent Support Program (BX2023375), Zhejiang Provincial Research Center for Upper Gastrointestinal Tract Cancer (JBZX-202006), Medical Science and Technology Project of Zhejiang Province (WKJ-ZJ-2202, WKJ-ZJ-2104), National Natural Science Foundation of China (82304946, 82074245, 81973634, 81903842, 81502603), Natural Science Foundation of Zhejiang Province (LR21H280001, TGY23H160038), Science and Technology Projects of Zhejiang Province (2019C03049, 2022KY684), Program of Zhejiang Provincial TCM Sci-tech Plan (2018ZY006, 2020ZZ005, 2022ZQ020), China Postdoctoral Science Foundation (2023M733563), and The Medicine and Health Science Fund of Zhejiang Province (2023KY073).

Author contribution

X.D.C., Z.Y.X., and L.S.: conceived and designed the analysis; M.X.C., C.H., and F.L.: contributed data or analysis tools; M.X.C., C.H., and F.L.: performed the analysis; M.X.C. and C.H.: wrote the paper. All authors contributed in collecting the data.

Conflicts of interest disclosure

Conflict of interest relevant to this article was not reported.

Research registration unique identifying number (UIN)

NCT05617469.

Guarantor

Xiangdong Cheng, Department of Gastric surgery, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, People's Republic of China. E-mail: chengxd@zjcc.org.cn.

Data availability statement

Due to the privacy of patients, the data related to patients cannot be available for public access but can be obtained from the corresponding author on reasonable request approved by the institutional review board of all enrolled centers.

Provenance and peer review

Submitted without invitation.

Presentation

None.

Acknowledgement

Assistance with the study: none.

References

[1] Thrift AP, El-Serag HB. Burden of gastric cancer. Clin Gastroenterol Hepatol 2020;18:534–42.
[2] Johnston FM, Beckman M. Updates on management of gastric cancer. Curr Oncol Rep 2019;21:67.
[3] Bray F, Ferlay J, Soerjomataram I, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.
[4] Joshi SS, Badgwell BD. Current treatment and recent progress in gastric cancer. CA Cancer J Clin 2021;71:264–79.
[5] Smyth EC, Nilsson M, Grabsch HI, *et al*. Gastric cancer. Lancet (London, England) 2020;396:635–48.
[6] Gadde R, Tamariz L, Hanna M, *et al*. Metastatic gastric cancer (MGC) patients: Can we improve survival by metastasectomy? A systematic review and meta-analysis. J Surg Oncol 2015;112:38–45.

[7] Liu D, Lu M, Li J, *et al*. The patterns and timing of recurrence after curative resection for gastric cancer in China. World J Surg Oncol 2016;14:305.
[8] Zhang ZY, Ge HY. Micrometastasis in gastric cancer. Cancer Lett 2013; 336:34–45.
[9] Mranda GM, Xue Y, Zhou XG, *et al*. Revisiting the 8th AJCC system for gastric cancer: a review on validations, nomograms, lymph nodes impact, and proposed modifications. Ann Med Surg (Lond) 2022;75:103411.
[10] Zhu Y, Fang X, Wang L, *et al*. A predictive nomogram for early death of metastatic gastric cancer: a retrospective study in the SEER database and China. J Cancer 2020;11:5527–35.
[11] Spolverato G, Capelli G, Lorenzoni G, *et al*. Development of a prognostic nomogram and nomogram software application tool to predict overall survival and disease-free survival after curative-intent gastrectomy for gastric cancer. Ann Surg Oncol 2022;29:1220–9.
[12] Zhao S, Bi Y, Wang Z, *et al*. Accuracy evaluation of combining gastroscopy, multi-slice spiral CT, Her-2, and tumor markers in gastric cancer staging diagnosis. World J Surg Oncol 2022;20:152.
[13] Chen Q, Zhang L, Liu S, *et al*. Radiomics in precision medicine for gastric cancer: opportunities and challenges. Eur Radiol 2022;32:5852–68.
[14] Wang R, Dai W, Gong J, *et al*. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. J Hematol Oncol 2022;15:11.
[15] Jiang Y, Liang X, Wang W, *et al*. Noninvasive prediction of occult peritoneal metastasis in gastric cancer using deep learning. JAMA Network Open 2021;4:e2032269.
[16] Gao X, Ma T, Bai S, *et al*. A CT-based radiomics signature for evaluating tumor infiltrating Treg cells and outcome prediction of gastric cancer. Ann Transl Med 2020;8:469.
[17] Zhang J, Cui Y, Wei K, *et al*. Deep learning predicts resistance to neoadjuvant chemotherapy for locally advanced gastric cancer: a multicenter study. Gastric Cancer 2022;25:1050–9.
[18] Mathew G, Agha R, Albrecht J, *et al*. STROCSS 2021: strengthening the reporting of cohort, cross-sectional and case-control studies in surgery. Int J Surg 2021;96:106165.
[19] Brendlin AS, Peisen F, Almansour H, *et al*. A machine learning model trained on dual-energy CT radiomics significantly improves immunotherapy response prediction for patients with stage IV melanoma. J Immunother Cancer 2021;911:e003261.
[20] Hu C, Chen W, Li F, *et al*. Deep learning radio-clinical signatures for predicting neoadjuvant chemotherapy response and prognosis from pretreatment CT images of locally advanced gastric cancer patients. Int J Surg 2023;109:1980–92.
[21] Sharma P, Hassan C. Artificial intelligence and deep learning for upper gastrointestinal neoplasia. Gastroenterology 2022;162:1056–66.
[22] Shimizu H, Nakayama KI. Artificial intelligence in oncology. Cancer Sci 2020;111:1452–60.
[23] Tran KA, Kondrashova O, Bradley A, *et al*. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med 2021;13:152.
[24] Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and deep learning in lung cancer. Strahlenther Onkol 2020;196:879–87.
[25] Liu X, Zhang D, Liu Z, *et al*. Deep learning radiomics-based prediction of distant metastasis in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy: a multicentre study. EBioMedicine 2021;69:103442.
[26] Li J, Dong D, Fang M, *et al*. Dual-energy CT-based deep learning radiomics can improve lymph node metastasis risk prediction for gastric cancer. Eur Radiol 2020;30:2324–33.
[27] Cui Y, Zhang J, Li Z, *et al*. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: a multicenter cohort study. EClinicalMedicine 2022;46:101348.
[28] 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 Digital Health 2022;4:e340–50.
[29] Zhang W, Fang M, Dong D, *et al*. Development and validation of a CT-based radiomic nomogram for preoperative prediction of early recurrence in advanced gastric cancer. Radiother Oncol 2020;145:13–20.
[30] Coccolini F, Nardi M, Montori G, *et al*. Neoadjuvant chemotherapy in advanced gastric and esophago-gastric cancer. Meta-analysis of randomized trials. Int J Surg 2018;51:120–7.
[31] Sah BK, Zhang B, Zhang H, *et al*. Neoadjuvant FLOT versus SOX phase II randomized clinical trial for patients with locally advanced gastric cancer. Nat Commun 2020;11:6093.
[32] Xue K, Ying X, Bu Z, *et al*. Oxaliplatin plus S-1 or capecitabine as neoadjuvant or adjuvant chemotherapy for locally advanced gastric cancer with D2 lymphadenectomy: 5-year follow-up results of a phase II-III randomized trial. Chin J Cancer Res 2018;30:516–25.