EVIDENCE REVIEW

6314 COVID-19 drug-induced liver injury: A recent update of the literature

Saha L, Vij S, Rawat K

REVIEW

6328 Gastrointestinal microbiota: A predictor of COVID-19 severity?

Neag MA, Vulturar DM, Gherman D, Burlacu CC, Todea DA, Buzoianu AD

MINIREVIEWS

6345 The mononuclear phagocyte system in hepatocellular carcinoma

Qiao DR, Shan GY, Wang S, Cheng JY, Yan WQ, Li HJ

6356 Rifabutin as salvage therapy for Helicobacter pylori eradication: Cornerstones and novelties


6363 Deep learning based radiomics for gastrointestinal cancer diagnosis and treatment: A minireview


ORIGINAL ARTICLE

Observational Study

6380 Best therapy for the easiest to treat hepatitis C virus genotype 1b-infected patients


Randomized Clinical Trial

6397 Endoscopic mucosal resection-precutting vs conventional endoscopic mucosal resection for sessile colorectal polyps sized 10-20 mm

Zhang XQ, Sang JZ, Xu L, Mao XL, Li B, Zhu WL, Yang XY, Yu CH

META-ANALYSIS

6410 Meta-analysis on the epidemiology of gastroesophageal reflux disease in China

Lu TL, Li SR, Zhang JM, Chen CW

CASE REPORT

6421 Germline BRCA2 variants in advanced pancreatic acinar cell carcinoma: A case report and review of literature

MINIREVIEWS

Deep learning based radiomics for gastrointestinal cancer diagnosis and treatment: A minireview

Pak Kin Wong, In Neng Chan, Hao-Ming Yan, Shan Gao, Chi Hong Wong, Tao Yan, Liang Yao, Ying Hu, Zhong-Ren Wang, Hon Ho Yu

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): B, B
Grade C (Good): 0
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Gnetti L, Italy; Litvin A, Russia

Received: September 20, 2022
Peer-review started: September 20, 2022
First decision: October 18, 2022
Revised: October 25, 2022
Accepted: November 16, 2022
Article in press: November 16, 2022
Published online: December 7, 2022

Abstract
Gastrointestinal (GI) cancers are the major cause of cancer-related mortality globally. Medical imaging is an important auxiliary means for the diagnosis, assessment and prognostic prediction of GI cancers. Radiomics is an emerging and effective technology to decipher the encoded information within medical images, and traditional machine learning is the most commonly used tool. Recent advances in deep learning technology have further promoted the development of radiomics. In the field of GI cancer, although there are several surveys on radiomics, there is no specific review on the application of deep-learning-based radiomics (DLR). In this review, a search was conducted on Web of Science, PubMed, and Google Scholar with an emphasis on the application of DLR for GI cancers, including esophageal, gastric, liver, pancreatic, and colorectal cancers. Besides, the challenges and recommendations based on the findings of the review are comprehensively analyzed to advance DLR.
INTRODUCTION

Gastrointestinal (GI) cancers, mainly include colorectal, gastric, liver, esophageal, and pancreatic cancers, and are the leading cause of cancer-related mortality globally[1]. According to CANCER TOMORROW[2], a forecast of the global burden of cancer mortality and incidence, by 2040, new cases of GI cancer and deaths will increase significantly. In recent years, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound (US) and other medical imaging techniques have been widely used in GI cancer diagnosis and treatment[3,4]. It is foreseeable that with the increase in GI cancer, the amount of medical imaging data will continue to grow. However, manual reading cannot cope with this growth, and the disparity in expertise among radiologists causes a high rate of missed diagnosis and misdiagnosis. In addition, traditional CT, MRI, PET, US, and other imaging examinations cannot observe changes in tumor heterogeneity, which can provide a better understanding of the causes and progression of cancer[5]. The development of radiomics technology provides new opportunities and methods to solve these dilemmas.

Radiomics is an emerging method for quantitative analysis and prediction of tumor phenotypes using machine learning or statistical models, and was proposed by Lambin et al[6] in 2012. In recent years, radiomics has been widely used in GI cancer and showed notable outcomes in tumor characterization, therapy response assessment, and prediction of survival rate after surgery[7-11]. Compared with the conventional method of using only manual inspection, radiomics can extract high-dimensional features that are difficult to be quantitatively described by the doctors from massive radiological images, and to correlate them with clinical and pathological data of patients in order to improve diagnosis and prognostication[12]. The fundamental premise of radiomics is that the developed descriptive models may produce useful prognostic, predictive and diagnostic information. Radiomics can be divided into two main categories: conventional radiomics, also referred to handcrafted radiomics (HCR) and deep-learning-based radiomics (DLR), also referred to as discovery radiomics[13]. Given the benefits of these two approaches, hybrid solutions that mix HCR and DLR also exist.

The HCR workflow is divided into multiple steps: (1) Image acquisition and reconstruction; (2) image segmentation and delineation of region of interest (automatic, semi-automatic, or manual delineation); (3) feature extraction and quantification. This is the core step of HCRs. The extracted features are mainly handcrafted features (also referred as pre-designed features), including shape, texture and intensity features. Some features may be highly correlated or redundant, so feature dimensionality reduction is an important step in feature analysis; and (4) Clinical target-oriented model building and validation. At this step, classic machine learning algorithms are usually used to develop high-precision and high-efficiency prediction models, and the models are trained and validated with sufficient data. The workflow of HCR is depicted in Figure 1.

Although HCR has been widely adopted in GI cancer and has achieved significant results, it has some deficiencies, such as low degree of automation and standardization, cumbersome and time-consuming feature extraction steps, and insufficient robustness and accuracy. Recently, deep learning, a promising technique in characterization of medical images, has gained much attention[14-17]. Many researchers have adopted DLR to overcome the limitations of conventional radiomics[18-22]. DL refers to a broad class of algorithms rather than a specific model. As long as a deep neural network structure is used to represent features at a deeper level, it can be called DL model. One of the most popular DL models used in medical imaging is convolutional neural networks (CNNs), which can automatically learn representative features from medical images. The use of CNNs in radiomics makes it easy to build an end-to-end feature extraction process, thereby avoiding the tedious and handcrafted feature extraction process. CNNs can also be used in image reconstruction and segmentation to improve the automation level of...
HCR, and the accuracy and reliability of diagnosis and prediction (Figure 1).

DL techniques are revolutionizing radiomics. In the field of GI cancer diagnosis and treatment, while there are several surveys on HCR[9-11], there is no specific review on the application of DLR. To provide a comprehensive overview of DLR in GI cancer, the performance of DLR in gastroenterology is summarized in this review, with an emphasis on the diagnosis and treatment of GI cancers, including esophageal, gastric, liver, pancreatic, and colorectal cancers. The original contributions to knowledge of this review are: (1) A unique interdisciplinary viewpoint on radiomics by discussing state-of-the-art DLR solutions; and (2) the challenges and recommendations based on the findings of the review are thoroughly analyzed to advance the field.

**DLR FOR ESOPHAGEAL CANCER**

Esophageal cancer is the seventh most prevalent form of cancer and the sixth most lethal cancer globally [1], and it is classified into esophageal squamous cell carcinoma (ESCC) or esophageal adenocarcinoma according to the type of cells. In consideration of the low overall 5-year survival rate of patients and the variation in responsiveness of patients to the current treatments such as neoadjuvant chemotherapy (NAC) and neoadjuvant chemoradiotherapy (NCRT) due to tumor heterogeneity, it is vital to have accurate diagnosis, pretreatment evaluation and survival rate prediction. The number of DLR studies regarding esophageal cancer has been growing, with most of the studies exploring treatment response, and the others investigating disease classification and survival rate prediction.
An important preoperative topic of esophageal cancer is diagnosis, yet the number of relevant DLR studies for diagnosis is minimal. Takeuchi et al.\textsuperscript{23} fine-tuned VGG16 to develop a DLR model for the diagnosis of esophageal cancer from CT scans, and its performance was comparable to that of the radiologists during testing, with a higher accuracy of 84.2% and specificity of 90.0%.

Response to treatment, especially NAC and NCRT, is one of the most popular research interests in the field of esophageal cancer. Hu et al.\textsuperscript{24} designed a CT-based model to predict the pathological complete response to NCRT of patients with ESCC using DL features, in which the support vector machine (SVM) classifier executed the classification action. The DL features were extracted using pretrained models, and the optimal one used ResNet50 that achieved an area under the receiver operating characteristic curve (AUC) of 0.805 and accuracy of 77.1% for the testing cohort, which achieved better results than using handcrafted features. Ypsilantis et al.\textsuperscript{25} designed a 36-CNN model that extracted DL features from PET scans and predicted whether the patient with esophageal cancer was non-responsive to NAC. This model was also compared with other competitive machine learning algorithms and results showed that it surpassed the other models with an average specificity, sensitivity and accuracy of 80.7%, 81.6%, and 73.4% respectively. Amyar et al.\textsuperscript{26} presented a novel 3D CNN model named 3D RPET-NET that predicted the response to CRT using esophageal cancer images of FDG-PET scans, and a comparative analysis with other approaches in the literature was also carried out. Three-dimensional RPET-NET obtained the best results with an accuracy of around 72% and even reached 75% when using tumor volume with an isotropic margin of 2 cm. Li et al.\textsuperscript{27} proposed a CT-based 3D DLR model (3D-DLRM), which was modified from ResNet34. Its aim was to predict whether patients with locally advanced thoracic ESCC had an objective or nonobjective response to concurrent CRT, achieving a validation AUC and positive predictive value of 0.833 and 100%, respectively. They also evaluated a model integrating the 3D-DLRM with clinical selected factors that even outperformed the individual 3D-DLRM, reaching a validation AUC of 0.861.

Other research interests of esophageal cancer include patient survival rate prediction. Wang et al.\textsuperscript{28} compared the use of an HCR model, DLR model and DLR nomogram for the prediction of the survival rate of esophageal cancer patients after 3 years of CRT, in which DL features were extracted and selected by DenseNet-169 to build the DLR model. This DLR nomogram attained the highest validation AUC of 0.942 and Harrell’s concordance index (C-index) of 0.784, surpassing the results produced by the sole use of HCR and DLR models. Yang et al.\textsuperscript{29} proposed a 3D-CNN model based on ResNet18 to predict esophageal cancer patient survival rate using PET scans. The model was initially pretrained to classify abnormal and healthy esophagus, and then trained to classify whether patients survived or expired within a year after diagnosis in the second stage, and the model obtained an AUC of 0.738. Gong et al.\textsuperscript{30} developed a hybrid radiomics nomogram to predict local recurrence-free survival (LRFS) of locally advanced ESCC patients who received definitive CRT from contrast-enhanced CT (CECT) scans, and it was combined with radiomic features, features extracted by 3D-DenseNet and prognostic clinical risk factors. The final model achieved a C-index of 0.76 for its external validation set, indicating the effectiveness of the addition of DL features for better prediction performances.

Some studies also discuss the application of DLR to prediction of lymph node (LN) metastasis, which is an effective prognosis factor of ESCC. Wu et al.\textsuperscript{31} built a model involving HCR, computer vision and DLR to predict the LN status of ESCC patients, and they also constructed two simpler models for efficacy comparison, and they exploited Convolution Neural Network-Fast (CNN-F) to extract DL features from CT images. The model with all signatures involved performed the best with C-statistic of 0.875, 0.874, and 0.840 for training, internal validation, and external validation cohorts, and those demonstrate its satisfactory discriminatory ability.

The studies about the application of DLR for esophageal cancer are summarized in Table 1.

### DLR FOR GASTRIC CANCER

Gastric cancer (GC) is the fifth most prevalent form of cancer and the fourth most lethal cancer globally\textsuperscript{1}. To ameliorate the low survival rate of patients, early diagnosis of disease and systematic treatment methods are necessary. The application of DLR in GC has been a promising area for research with a rising number of relevant studies published every year, that aim to tackle or refine the existing concerns regarding GC.

Many studies focused on prediction of treatment response of patients. Cui et al.\textsuperscript{32} constructed a pretreatment venous-phase CT-based DLR nomogram that combined handcrafted features, DL features and remarkable clinicopathological factors to identify locally advanced GC patients with good response to NAC. The nomogram achieved better than the clinical model and the separate use of two features that were built for comparison, attaining C-index values of 0.829, 0.804, and 0.827 in its internal validation cohort and two external validation cohorts, respectively. Li et al.\textsuperscript{33} developed a combined artificial intelligence (AI) model that incorporated feature outputs from HCR and DLR models, which aimed to determine whether the patients had signet ring cell carcinoma (SRC) of GC and predict survival and treatment response to postoperative chemotherapy from CECT images. They also compared its efficacy with the clinical, HCR and DLR models, and the AI model obtained the best...
results with an AUC of 0.786 and accuracy of 71.6% for diagnosing SRCC for the test cohort. The AI model also evaluated that SRCC patients with higher risks had shorter median overall survival (OS) and insignificant improvements in median OS after receiving adjuvant chemotherapy than those of lower risk, indicating its good capability to predict survival and response to treatment. Tan et al.[34] built a dual-energy CT delta radiomics model to predict the treatment response to chemotherapy of patients with far-advanced GC. They developed a V-Net segmentation model, and the application of this semi-automatic segmentation model to the delta radiomics model shortened the diagnostic time and achieved better results in terms of mean AUC (0.728 vs 0.687 in the testing cohort, 0.828 vs 0.749 in the independent validation cohort) than using manual segmentation.

Survival rate prediction is also a popular topic for DLR of GC. Hao et al.[35] combined clinical variables, radiomic features and DL features to build a CT-based prediction Cox proportional-hazard model, which served to predict the OS and progression-free survival (PFS) of patients with GC. The model acquired the highest C-index of 0.783 and 0.770 for OS and PFS when using postoperative clinical variables, and the most dominant variables for survival prediction were identified as important prognostic factors in the subsequent survival analysis. Some studies only made use of DL techniques to build predictive models for similar purposes. Zhang et al.[36] proposed a multi-focus and multi-level fusion feature pyramid network (MMF-FPN) to predict OS risks of GC patients from CT images, and other models using existing methods in the literature were used for comparison. The experimental results showed that MMF-FPN was the finest model that attained the highest C-indexes (validation: 0.833; PPV: 100% vs 0.749 in the testing cohort) than using manual segmentation.

Table 1 Summary of studies using deep-learning-based radiomics for esophageal cancer

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Imaging</th>
<th>Study design</th>
<th>Study aim</th>
<th>DL model</th>
<th>Dataset</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takeuchi et al [23], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Detection of esophageal cancer</td>
<td>VGG16</td>
<td>1646 CT images (1500 images for training and validation, 146 for testing)</td>
<td>Accuracy: 84.2%; F value: 74.2%; Sensitivity: 71.1%; Specificity: 90% in test set</td>
</tr>
<tr>
<td>Hu et al[24], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Evaluation of response to NCRT to ESCC</td>
<td>ResNet50</td>
<td>231 patients (161 in training cohort, 70 in testing cohort)</td>
<td>AUC: 0.805; C-index: 0.805; Accuracy: 77.1%; Sensitivity: 83.9%; Specificity: 71.8% for the testing cohort</td>
</tr>
<tr>
<td>Ypsilantis et al [25], 2015</td>
<td>PET</td>
<td>Retrospective</td>
<td>Prediction of response to NAC in patients with esophageal cancer</td>
<td>3S-CNN</td>
<td>107 patients</td>
<td>Sensitivity: 80.7%; Specificity: 81.6%; Accuracy: 73.4%</td>
</tr>
<tr>
<td>Amyar et al [26], 2019</td>
<td>PET</td>
<td>Retrospective</td>
<td>Prediction of response to radio-chemotherapy in patients with esophageal cancer</td>
<td>3D RPET-NET</td>
<td>97 patients</td>
<td>Accuracy: 75.0%; Sensitivity: 76.0%; Specificity: 74.0%; AUC: 0.74</td>
</tr>
<tr>
<td>Li et al[27], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of treatment response to CCRT among patients with locally advanced TESCC</td>
<td>ResNet34</td>
<td>306 patients (203 in training cohort, 103 in validation cohort)</td>
<td>AUC: 0.833; PPV: 100%</td>
</tr>
<tr>
<td>Wang et al[28], 2022</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of survival rates for patients with esophageal cancer after 3 yr with chemoradiotherapy</td>
<td>DenseNet-169</td>
<td>154 patients (116 in training cohort, 38 in validation cohort)</td>
<td>AUC: 0.942; C-index: 0.784</td>
</tr>
<tr>
<td>Yang et al[29], 2019</td>
<td>PET</td>
<td>Retrospective</td>
<td>Identification of esophageal cancer patients with poor prognosis</td>
<td>3D-CNN based on ResNet18</td>
<td>1107 scans</td>
<td>AUC: 0.738</td>
</tr>
<tr>
<td>Gong et al[30], 2022</td>
<td>CECT</td>
<td>Retrospective</td>
<td>Prediction of LRFS in esophageal cancer patients after 1 yr of definitive chemoradiotherapy</td>
<td>3D-Densenet</td>
<td>397 patients</td>
<td>C-index: 0.76</td>
</tr>
<tr>
<td>Wu et al[31], 2019</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of LN status of patients with ESCC</td>
<td>CNN-F</td>
<td>411 patients</td>
<td>C-index: 0.840</td>
</tr>
</tbody>
</table>

CT: Computed tomography; PET: Positron emission tomography; CECT: Contrast-enhanced computed tomography; NCRT: Neoadjuvant chemoradiotherapy; NAC: Neoadjuvant chemotherapy; ESCC: Esophageal squamous cell carcinoma; CCRT: Concurrent chemo-radiation therapy; TESCC: Thoracic esophageal squamous cell carcinoma; LRFS: Local recurrence-free survival; LN: Lymph node; CNN-F: Convolutional neural network-fast; AUC: Area under the receiver operating characteristic curve; PPV: Positive predictive value; DLR: Deep-learning-based radiomics.
and inter-reader agreement, exceeding the results obtained by the radiomics signature and clinical modality built for comparison.

Accurate prediction of LN status of GC, which is a remarkable prognostic factor, is of importance to determine the appropriate treatment. Guan et al.[38] explored the efficacy of using different DL models to extract features and machine learning classifiers (i.e., SVM and random forest) to build a CT-based predictive model for the evaluation of LN status. Other models using radiomic features and integrated features were built for comparison, and the best model was ResNet50-RF with an AUC and accuracy of 0.9803 and 98.10%, respectively. A nomogram based on DL feature scores and clinical risk factors was also developed and a higher AUC of 0.9914 was achieved in the testing cohort. Dong et al.[39] proposed a similar DLR nomogram to evaluate the number of LN metastases of locally advanced GC patients before surgery, in which radiomics signatures that contained handcrafted and DL features and clinical characteristics were used. The performance of the model was evaluated with four validations sets; three of which were collected from China and one from Italy. The model showed its good discriminative capability to identify N-staging of GC with higher C-indexes of 0.797 in the validation sets from China and 0.822 in the set from Italy, and it outperformed other predictors such as clinical models and single signatures. To predict the LN status and prognosis of patients, the dual-energy CT-based DLR nomogram created by Li et al.[40] incorporated CT-reported LN and two radiomics signatures for arterial-phase and venous-phase CT images, in which DL features were extracted via CNN. The nomogram performed better and gained a higher AUC of 0.82 than the clinical model built alongside for comparative analysis, and the associated prognostic prediction was satisfactory in terms of PFS (C-index: 0.64) and OS (C-index: 0.67). Jin et al.[41] developed a DLR model that adopted ResNet-18 to evaluate the LN status in nodal stations using CECT, and the high value of the median AUC of the 11 stations (0.876) proved the excellent prediction ability of the model. The authors attempted to build a nomogram combining the DL features with clinical features, but no significant improvements in the results were observed.

Other prognostic factors of GC have also been investigated in previous studies. Sun et al.[42] exploited DL techniques to build a CT-based radiomics nomogram for evaluating the status of serosal invasion of advanced GC patients. Three radiomics signatures were generated based on the three phases of CT images with their DL features extracted using CNNs, and they were integrated with clinical characteristics to form the nomogram. The final model outperformed other models, such as clinical and phenotypic models, and its AUC for test sets I and II was 0.87 and 0.90, respectively. Li et al.[43] compared the use of DL features and radiomic features to create a CECT-based GC risk (GRISK) model using similar procedures for the prediction of the status of lymphovascular invasion in patients with localized GC. The team explored the use of deep transfer learning models to build a gastric imaging marker, in which five pretrained models and an auto-encoder were utilized for feature extraction and reduction, respectively. Then, it was integrated with patient clinical and radiological characteristics to construct its own GRISK model. The GRISK model with deep transfer learning gastric imaging marker obtained comparable AUC (0.722 vs 0.725) and accuracy (0.671 vs 0.710) with the other model with the radiomics gastric imaging marker but did not surpass the latter model.

The studies investigating the usage of DLR for GC are summarized in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients</th>
<th>Model</th>
<th>Validation</th>
<th>AUC</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guan et al.</td>
<td>38</td>
<td>ResNet50-RF</td>
<td>Validation</td>
<td>0.9803</td>
<td>98.10%</td>
</tr>
<tr>
<td>Dong et al.</td>
<td>39</td>
<td>ResNet50-RF</td>
<td>Validation</td>
<td>0.822</td>
<td></td>
</tr>
<tr>
<td>Li et al.</td>
<td>40</td>
<td>ResNet50-RF</td>
<td>Validation</td>
<td>0.876</td>
<td></td>
</tr>
<tr>
<td>Jin et al.</td>
<td>41</td>
<td>ResNet-18</td>
<td>Validation</td>
<td>0.876</td>
<td></td>
</tr>
<tr>
<td>Sun et al.</td>
<td>42</td>
<td>ResNet-18</td>
<td>Validation</td>
<td>0.87 and 0.90</td>
<td></td>
</tr>
<tr>
<td>Li et al.</td>
<td>43</td>
<td>ResNet-18</td>
<td>Validation</td>
<td>0.876</td>
<td></td>
</tr>
</tbody>
</table>

**DLRs FOR LIVER CANCER**

Primary liver cancer is the sixth most prevalent form of cancer and the third most lethal cancer globally, and some of its common phenotypes are hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma[1]. Taking the high mortality caused by this disease into account, the clinical application of early diagnosis, individualized evaluation and prognosis prediction are valued. The exploitation of DLR technology in liver cancer has been rapidly developing, and various solutions for the issues in different phases of diagnosis and treatment are emerging.

Computer-aided diagnosis does not only aid radiologists such as shortening the diagnosis time, but also allow them to evaluate appropriate treatments at earlier stages of liver cancer. Ding et al.[44] constructed a CT-based DLR model that fused a radiomics signature and a DL model, to differentiate HCC into low or high grade. The DL model was an alteration of VGG19 and it performed better than the radiomics signature, with better AUC (0.7513 vs 0.7475) and accuracy (66.31% vs 65.78%). The fused DLR model was the optimal model with observable improvements in the results, achieving an AUC of 0.8042 and accuracy of 72.73%.

Accurate prediction of patient response to different therapies is critical to realize personalized treatment at different stages of HCC. Peng et al.[45] developed a multi-class DL model from ResNet50 to predict four treatment responses to transarterial chemoembolization (TACE) therapy of HCC patients using CECT scans. Its performance was assessed using confusion matrices and receiver operating characteristic curves, and the model attained an AUC over 0.90 for all four classes in both validation sets, and accuracies of 85.1% and 82.8% for validation sets 1 and 2, respectively. In the next year, they combined conventional radiomics and DL to build a new CECT-based DLR model that served to predict the initial treatment response to TACE of HCC patients preoperatively[46]. Different from their prior...
prediction of response to chemotherapy in patients with gastric cancer

Mean AUC: 0.728 (testing)

Prediction of OS risks of AGC patients

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of lymphovascular invasion of localized gastric cancer patients

C-index: 0.829

CT: Computed tomography; NAC: Neoadjuvant chemotherapy; LAGC: Locally advanced gastric cancer; SRCC: Signet ring cell carcinoma; OS: Overall survival; PFS: Progression-free survival; AGC: Advanced gastric cancer; LNM: Lymph node metastasis; LN: Lymph node; DCNNs: Deep convolutional neural networks; MMN-FPN: Multi-focus and multi-level fusion feature pyramid network; UC: Area under the receiver operating characteristic curve; C-index: Harrell’s concordance index; GRISK: Gastric Risk; RF: Random forest; VAE: Variational auto-encoder; DLR: Deep-learning-based radiomics.

Prediction of response to chemotherapy in AGC patients

AUC: 0.82 in test set; C-index of OS: 0.67; C-index of PFS: 0.64

Prediction of the number of LNM in AGC

Median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort

Evaluation of lymphovascular invasion of localized gastric cancer patients

AUC: 0.725; sensitivity: 73.2%; specificity: 63.3%; accuracy: 67.0% for radiomics GRISK model (final model) in testing cohort

stage II: Validation

AUC: 0.87; sensitivity: 80%; specificity: 73%; accuracy: 85%; median AUC: 0.85 in test set I. AUC: 0.90; accuracy: 85%; specificity: 75%; sensitivity: 93 in test set II

SqueezeNet, ResNet50, Inception V3, VGG19, DeepLoc

Prediction of response to NAC in patients with LAGC

C-index: 0.822 in validation set

Prediction of the number of LN stations of gastric cancer patients

Median AUC: 0.828 (validation set); 0.828 (validation cohort) when using semi-segmentation

Attention-guided VAE

C-index of OS: 0.783; C-index of PFS: 0.770 when only using postoperative variables

Prediction of preoperative status of LNM of gastric cancer patients

AUC: 0.9803; accuracy: 98.10%; sensitivity: 98.39%; specificity: 0.9676% for testing of ResNet50-RF

Prediction of OS after gastrectomy; evaluation of effects of variables on survival prediction

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of OS risks of patients with gastric cancer

AUC: 0.80; sensitivity: 0.73; specificity: 0.667

Prediction of early recurrence of patients with AGC

AUC: 0.82; sensitivity: 0.73; specificity: 0.667

Prediction of OS and PFS after gastrectomy; evaluation of effects of variables on survival prediction

AUC: 0.725; sensitivity: 73.2%; specificity: 0.93 in test set

AUC: 0.9803; accuracy: 98.10%; sensitivity: 98.39%; specificity: 0.9676% for testing of ResNet50-RF

Sensitivity: 0.743; median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort

Sensitivity: 0.743; median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort

AUC: 0.82; sensitivity: 0.73; specificity: 0.667

AUC: 0.87; sensitivity: 80%; specificity: 73%; accuracy: 85%; median AUC: 0.85 in test set I. AUC: 0.90; accuracy: 85%; specificity: 75%; sensitivity: 93 in test set II

SqueezeNet, ResNet50, Inception V3, VGG19, DeepLoc

AUC: 0.725; sensitivity: 73.2%; specificity: 60.3%; accuracy: 67.0% for radiomics GRISK model (final model) in testing cohort

Prediction of response to chemotherapy in patients with gastric cancer

AUC: 0.728 (testing)

Prediction of response to chemotherapy in patients with gastric cancer

AUC: 0.82 in test set; C-index of OS: 0.67; C-index of PFS: 0.64

Attention-guided VAE

C-index: 0.822 in validation set

Prediction of response to chemotherapy in AGC patients

AUC: 0.82; sensitivity: 0.73; specificity: 0.667

Prediction of the number of LNM in LAGC

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of LNM and prognosis in gastric cancer patients

AUC: 0.82 in test set; C-index of OS: 0.67; C-index of PFS: 0.64

Prediction of OS of patients with gastric cancer

AUC: 0.9803; accuracy: 98.10%; sensitivity: 98.39%; specificity: 0.9676% for testing of ResNet50-RF

MMF-FPN

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of early recurrence of patients with AGC

AUC: 0.80; sensitivity: 0.73; specificity: 0.667

Prediction of OS of patients with gastric cancer

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of OS and PFS after gastrectomy; evaluation of effects of variables on survival prediction

AUC: 0.725; sensitivity: 73.2%; specificity: 0.93 in test set

Prediction of LNM status in LN stations of gastric cancer patients

Median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort

Prediction of OS of patients with gastric cancer

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of response to chemotherapy in AGC patients

AUC: 0.82; sensitivity: 0.73; specificity: 0.667

Prediction of early recurrence of patients with AGC

AUC: 0.82 in test set; C-index of OS: 0.67; C-index of PFS: 0.64

Prediction of OS of patients with gastric cancer

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of LNM in LAGC

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of OS risks of patients with gastric cancer

AUC: 0.725; sensitivity: 73.2%; specificity: 0.93 in test set

Prediction of OS after gastrectomy; evaluation of effects of variables on survival prediction

AUC: 0.82; sensitivity: 0.73; specificity: 0.667

Prediction of early recurrence of patients with AGC

AUC: 0.80; sensitivity: 0.73; specificity: 0.667

Prediction of OS of patients with gastric cancer

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of OS risks of patients with gastric cancer

AUC: 0.725; sensitivity: 73.2%; specificity: 0.93 in test set

Prediction of the number of LNM in LAGC

C-index: 0.76; hazard ratio: 9.46 in test set

Prediction of LNM status in LN stations of gastric cancer patients

Median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort

Prediction of LNM status in LN stations of gastric cancer patients

Median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort

Stage I: Retrospective; stage II: Validation

Median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort

Prediction of lymphovascular invasion of localized gastric cancer patients

C-index: 0.829

CT: Computed tomography; NAC: Neoadjuvant chemotherapy; LAGC: Locally advanced gastric cancer; SRCC: Signet ring cell carcinoma; OS: Overall survival; PFS: Progression-free survival; AGC: Advanced gastric cancer; LNM: Lymph node metastasis; LN: Lymph node; DCNNs: Deep convolutional neural networks; MMN-FPN: Multi-focus and multi-level fusion feature pyramid network; UC: Area under the receiver operating characteristic curve; C-index: Harrell’s concordance index; GRISK: Gastric Risk; RF: Random forest; VAE: Variational auto-encoder; DLR: Deep-learning-based radiomics.

work, they designed their own CNN for feature extraction and prediction, and the DL model was integrated with five radiomics models built with different classic machine learning algorithms or tumor size feature to build integrated models for efficacy comparison. The DL model outperformed all individual radiomics models with an AUC of 0.972, while all integrated models yielded higher values of AUC than merely using DL model. The combination of DL with random forest classifier obtained the highest AUC of 0.994.

Survival prediction is also an important research area to facilitate individualized HCC treatment. To predict the OS of HCC patients who were treated with stereotactic body radiation therapy, Wei et al.[47] established a CECT-based DL network model that comprised two variational-autoencoder-based survival models and one CNN-based model for extracting radiomic features, clinical features and CT features. The performance of the separate models and the integrated radiomics model using either DL network or Cox hazard model was compared by C-index, in which the integrated model produced the
highest C-index of 0.650 in repeated cross-validation among all models. Liu et al.\(^5\) developed two separate DLR models to differentiate HCC patients who received radiofrequency ablation (RFA) or surgical resection into high or low risks using CEUS images, and the corresponding radiomics signatures were built. Afterwards, two radiomics nomograms were constructed by combining the signatures with clinical variables to predict the 2-year PFS of patients and both models. Both DLR models achieved satisfactory values of C-index (0.726 for RFA, 0.741 for surgical resection). The good agreement of the survival predictions of the nomograms was demonstrated from the calibration curves.

Postoperative recurrence of cancer is one of the primary causes of death, which extends to the increase in recurrence risk assessment using DLR. To predict the early recurrence of HCC patients using multi-phase CECT scans, Wang et al.\(^49\) explored the predictive ability of various kinds of models, and they included a DLR model based on ResNet, a clinical model extracting features from clinical data and three combined CNN-based models of different structures. Experimental results demonstrated that the integration of DL features and clinical features improved the prediction accuracy, and one combined model obtained the highest AUC of 0.825. The team improved their study by comparing the DL model with a conventional radiomics model, and one more combined model of another structure was added to the comparative analysis of their previous work.\(^50\) The DL model performed better than the radiomics model with an average AUC of 0.7233 and accuracy of 69.52%, while one of the combined models surpassed the rest in the comparative analysis and reached 0.8248 and 78.66% in its average AUC and accuracy, respectively. They also investigated the effect of attaching a joint loss function to the best model on the average AUC and accuracy, and the two metrics were improved to 0.8331 and 80.49%. He et al.\(^51\) presented an intelligent-augmented DL model for Risk Assessment of Post Liver Transplantation (i-RAPIT) model in their study, which was a multi-network model that estimated the recurrence risk of HCC patients after liver transplantation. The i-RAPIT model was composed of two deep CapsNet networks for feature extraction from MR and pathological images, and a natural-language-processing-based radial basis function (NLP-based RBF) for extracting clinical features. Before the MR images were entered into the model, U-Net was also exploited for tumor and liver detection in the images. The model achieved a total accuracy of 82%, and AUC of 0.87 and F-1 score of 84% when comparing with other network combinations.

Early detection of microvascular and macrovascular invasion is another practical approach to select the proper therapy for HCC patients and reduce mortality. Jiang et al.\(^52\) adopted 3D-CNN to build a CT-based DL model for predicting the status of microvascular invasion of HCC patients, and three models based on radiomics features, radiologic features, and integration of the two kinds of features and clinical characteristics was also used for comparison. The results produced by the four models were excellent, with the DL model achieving better results for a few metrics such as AUC (0.906), and sensitivity (93.2%) in the validation set. Wang et al.\(^53\) devised a new DL model named MVI-Mind that consisted of a light-weight transformer for segmentation and a CNN for prediction of microvascular invasion, and several DL techniques were used to compare the proposed methods. The MVI-Mind attained highest mean intersection over union of 0.9006 and accuracy of 99.47% as compared with other DL segmentation algorithms, and it maintained its superiority in prediction and obtained AUC values of 0.9223, 0.8962, and 0.9100 for arterial phase, portal venous phase and delayed period CT images, respectively. For estimating the status of macrovascular invasion using CT scans, Fu et al.\(^54\) utilized the concept of multi-task DL neural network (MTnet) to build predictive models. Radiomic features from CT images, clinical and radiological factors were fused to construct the proposed model, and it was modified from U-Net that contained modules engaged in tumor segmentation, feature extraction and prediction. It exhibited the most outstanding performance with an AUC of 0.836 among all models built for comparison.

The studies investigating the implementation of DLR for liver cancer are summarized in Table 3.

---

**DLR FOR PANCREATIC CANCER**

Pancreatic cancer is the seventh most deadly cancer worldwide, in which pancreatic adenocarcinoma or pancreatic ductal adenocarcinoma (PDAC) are the most prevalent, accounting for the high mortality rate [1]. The number of deaths caused by this disease is almost equivalent to the number of cases due to the overall poor prognosis, so the introduction of advanced AI technologies is essential and urgent to rectify the situation. In these few years, the field of DLR in pancreatic cancer has flourished and more critical issues such as disease differentiation and survival prediction have been discussed.

Achieving an accurate diagnosis of PDAC gives a great contribution to avoiding false predictions and improving the survival outcomes of patients. For distinguishing between PDAC and autoimmune pancreatitis using CT scans, Ziegelmayer et al.\(^55\) developed a DLR model that utilized VGG19 to extract DL features, and its efficacy was compared with a model trained on handcrafted radiomic features. The former model performed better with higher mean values in AUC, sensitivity, and specificity (0.90, 89% and 83%) over the cross-validation procedure. Liao et al.\(^56\) used a DL model based on the coarse-to-fine network architecture search (C2FNAS) to perform segmentation of CECT images for radiomic feature extraction, and they were used for training the machine learning model for
Table 3 Summary of studies using deep-learning-based radiomics for liver cancer

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Imaging</th>
<th>Study design</th>
<th>Study aim</th>
<th>DL model</th>
<th>Dataset</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ding et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Evaluation of HCC differentiation</td>
<td>VGG19</td>
<td>1234 patients (799 in training cohort, 248 in validation cohort; 187 in independent testing cohort)</td>
<td>AUC: 0.8042; accuracy: 72.73%; sensitivity: 70.75%; specificity: 75.31% in testing cohort for the fused DLRs model</td>
</tr>
<tr>
<td>Peng et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of different treatment responses to TACE in HCC patients</td>
<td>ResNet50</td>
<td>789 patients (562 in training cohort; 89 and 136 in validation cohorts 1 and 2)</td>
<td>Accuracy: 85.1% in validation cohort 1; accuracy: 82.8% in validation cohort 2</td>
</tr>
<tr>
<td>Peng et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of initial response to TACE in HCC patients</td>
<td>CNN</td>
<td>310 patients (139 in training cohort; 171 in validation cohort)</td>
<td>AUC: 0.994</td>
</tr>
<tr>
<td>Wei et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of OS of HCC patients treated with SBRT</td>
<td>CNN</td>
<td>167 patients</td>
<td>C-index: 0.650 in cross validation</td>
</tr>
<tr>
<td>Liu et al</td>
<td>US</td>
<td>Retrospective</td>
<td>Prediction of PFS of HCC patients treated with RFA or surgical resection</td>
<td>CNN</td>
<td>214 RFA patients (149 for training; 65 for validation), 205 SB patients (144 for training; 61 for validation)</td>
<td>C-index of RFA: 0.726; C-index of surgical resection: 0.726</td>
</tr>
<tr>
<td>Wang et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of early recurrence of HCC patients</td>
<td>ResNet</td>
<td>167 patients</td>
<td>AUC of best model: 0.825</td>
</tr>
<tr>
<td>Wang et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of early recurrence of HCC patients</td>
<td>ResNet</td>
<td>167 patients</td>
<td>For the best model with joint loss function; AUC: 0.8313; accuracy: 80.49%</td>
</tr>
<tr>
<td>He et al</td>
<td>MRI and CT</td>
<td>Retrospective</td>
<td>Evaluation of HCC recurrence risk of liver transplantation recipients</td>
<td>U-net, CapsNet</td>
<td>109 patients (87 for training; 22 for testing)</td>
<td>Total accuracy: 82%; recall: 80%; precision: 89%; AUC: 0.87; F-1 score: 84%</td>
</tr>
<tr>
<td>Jiang et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of microvascular invasion status of HCC patients</td>
<td>3D-CNN</td>
<td>405 patients (324 in training set, 81 in validation set)</td>
<td>AUC: 0.906; sensitivity: 75.7%; specificity: 93.2%; accuracy: 85.2%; F-1 score: 87.2% in validation set</td>
</tr>
<tr>
<td>Wang et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of microvascular invasion status of HCC patients</td>
<td>Transformer, CNN</td>
<td>138 patients</td>
<td>For arterial phase images in validation set, AUC: 0.9223; Average accuracy: 86.78%</td>
</tr>
<tr>
<td>Fu et al</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of microvascular invasion status in HCC patients</td>
<td>Modified U-Net</td>
<td>366 patients (281 in training cohort, 85 in validation cohort)</td>
<td>AUC: 0.836 in validation cohort</td>
</tr>
</tbody>
</table>

CT: Computed tomography; MRI: Magnetic resonance imaging; US: Ultrasound; HCC: Hepatocellular carcinoma; TACE: Transarterial chemoembolization; PFS: Progression-free survival; RFA: Radiofrequency ablation; SBRT: Stereotactic body radiation therapy; CNN: Convolutional neural network; AUC: Area under the receiver operating characteristic curve; C-index: Harrell’s concordance index; DLR: Deep-learning-based radiomics.

Prediction. The DL segmentation model obtained a mean Dice score of 0.773 for segmentation while the prediction model yielded an AUC of 0.960 when distinguishing between PDAC and the control group (non-cancerous diseases and normal pancreas). Tong et al[57] constructed a ResNet-50-based DLRs model to classify PDAC and chronic pancreatitis patients from CEUS images, and the outputs were the probability of being PDAC or chronic pancreatitis, and heatmaps with highlighted regions that displayed the detected lesions. A two-round reader study was conducted to test the effectiveness of the model. The model achieved an AUC of 0.967 and 0.953 in two validation sets and outperformed the radiologists in the first round, while radiologists could obtain higher accuracies in determining the disease with the aid of the model in the second round.

Prediction of treatment response is also a critical aspect in the field of DLR in pancreatic cancer. Watson et al[58] built a CNN model based on LeNet to classify, using CT scans, whether PDAC patients had a pathological response or no response to NAC. It was compared with two models: a hybrid DL model that had the same architecture as the pure DL model but captured both CT image features and one clinical feature [≥ to 10% decrease in carbohydrate antigen (CA)-19], and a CA-19 model only taking in the feature regarding CA-19 decrease. Both DL models could produce superior results than the CA-19 model, and the hybrid DL model obtained a slightly higher AUC than the pure DL model (0.784 vs 0.738).

Survival prediction is another vital feature of PDAC that occupies a substantial portion of the existing DLR studies. Muhammad et al[59] designed a CNN architecture modified from AlexNet to evaluate the survival risk of PDAC patients that received radiomic features extracted from CECT images, and the
model reached a C-index of 0.85, indicating itself as a good survival model. Zhang et al[60] also made use of a CNN that was pretrained with non-small cell lung cancer images to construct their CT-based survival model for patients with resectable PDAC, in which a modified loss function was used. The proposed model accomplished finer prognostic predictions than the conventional radiomic model with an index of prediction accuracy of 11.81% and C-index of 0.651. They released another paper in the same year and compared the efficacy of DL and radiomic features from CECT images by feeding them separately to a random forest classifier to build a DLR model for predicting OS[61]. Similar to their prior work, the DL features were extracted by a pretrained CNN model but with a different structure. The model that used DL features attained an AUC of 0.81, which was higher than the other model based on radiomic features and gained a hazard ratio of 1.38 when the respective risk scores (predicted probabilities of deaths) were tested in survival analyses. Later, they further modified their previous DLR model to a risk score-based feature fusion model to predict 2-year OS[62]. Two small models based on DL and radiomic features separately were embodied in the framework to generate their corresponding risk scores, and these risk scores were used to train the main prediction model. The performance of the proposed model was later assessed with other models using different feature reduction techniques, and the risk score model achieved the highest AUC of 0.84. Yao et al[63] devised a new multi-task network model to perform both survival and tumor surgical margin prediction of resectable PDAC patients simultaneously using multi-phase CECT scans. Inside the model, a 3D-CNN model incorporated with a nnUNet for pancreas segmentation was exploited for the margin prediction part, while the combination of 3D-ResNet18 and Contrast-Enhanced 3D Convolutional Long Short-Term Memory (CE-ConvLSTM) network was responsible for survival prediction. The model achieved the results exceeding all other deep models in the comparative analysis, which yielded a C-index of 0.705 in predicting survival outcome and a balanced accuracy of 73.6% in determining the resection margin. They revised their preliminary work by incorporating pancreatic anatomical features into the model and switching to implement an automatically self-learning segmentation method that used 3D U-Net as the network architecture and nnUNet as the backbone model for training[63]. The new model attained the highest survival C-index of 0.667 and balanced accuracy of 67.1% for resection margin prediction among all the models including their previous model and other DL and radiomics models.

LN metastasis also possesses a high prognostic value in pancreatic cancer and it is noteworthy to have an early and accurate prediction of its status. An et al[64] developed a DLR model with different radiomics signatures extracted from dual-energy CT scans for the prediction of LN metastasis by a pretrained ResNet-18 model. Experiments of adding key clinical features were conducted to compare the effectiveness of using different approaches. The combined model integrated DL features and key clinical features yielded the highest AUC of 0.92 and accuracy of 86%.

The expression of various genes is an influential factor for patient prognosis and preoperative prediction of these prognostic factors can assist the diagnosis and treatment evaluation process. To predict the status of HMGA2 and C-MYC gene expression of PDAC patients, Li et al[65] compared the use of radiomic features, DL features (extracted by pretrained CNN) and integration of both features in a CT-based model using an SVM classifier. Region of interest segmentation was conducted by two experienced radiologists individually, and the model was tested with different segmented images for improving the validity of the study. A model using DL features and all features achieved similar values in all evaluation metrics for both C-MYC and HMGA2 tests, while DL features selected by Doctor B obtained outstanding average AUC scores (C-MYC: 0.90, HMGA2: 0.91) and accuracies (C-MYC: 95%, HMGA2: 88%) in the two gene tests.

The studies investigating the application of DLR for pancreatic cancer are summarized in Table 4.

### DLR FOR COLORECTAL CANCER

Colorectal cancer (CRC) is the third most common kind of cancer and the second leading cause of cancer-related fatalities worldwide[1]. It is crucial to carry out research on the diagnosis, treatment response prediction, and survival prediction of CRC, which can improve the prognosis of patients and significantly reduce the social and medical burden. In recent years, promising research results have emerged in the preoperative, intraoperative and postoperative stages of CRC using DLRs technology, covering the entire process of CRC diagnosis and treatment.

DLR is revolutionizing the treatment options for CRC. When making treatment decisions for CRC patients, identifying KRAS mutations, which may contribute to the continued proliferation of tumors, can help personalize treatment and care for CRC patients[66]. For preoperative prediction of KRAS mutations in patients with CRC, HCR and DLR were merged into a noninvasive model created by Wu et al[67]. The model, which mixed the handcrafted and DLR radiomics features, produced a C-index for the original cohort of 0.815 and the validation cohort of 0.832, which was higher than using HCR or DLR alone. For the individualized treatment decision-making in colorectal liver metastases (CRLM) management, the prediction of chemotherapeutic response is crucial. To predict the response to chemotherapy in CRLM, Wei et al[68] developed a ResNet10-based DLR model that used contrast-enhanced multidetector CT images as inputs. They also developed an HCR model for comparison.
## Table 4 Summary of studies using deep-learning-based radiomics for pancreatic cancer

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Imaging</th>
<th>Study design</th>
<th>Study aim</th>
<th>DL model</th>
<th>Dataset</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ziegelmayer et al[55], 2020</td>
<td>CT</td>
<td>Retrospective</td>
<td>Identification of PDAC and AIP</td>
<td>VGG19</td>
<td>86 patients (44 AIP patients and 42 PDAC patients)</td>
<td>Sensitivity: 89%; specificity: 83%; AUC: 0.90</td>
</tr>
<tr>
<td>Liao et al[56], 2022</td>
<td>CT</td>
<td>Retrospective</td>
<td>Identification of PDAC, non-cancerous pancreatic diseases and normal pancreas</td>
<td>CNN</td>
<td>3120 images (1872 for training, 624 for validation, 624 for testing)</td>
<td>Sensitivity: 89.9%; specificity: 91.3%; AUC: 0.960 when distinguishing between PDAC and control group</td>
</tr>
<tr>
<td>Tong et al[57], 2022</td>
<td>US</td>
<td>Retrospective</td>
<td>Identification of PDAC and CP</td>
<td>ResNet-50</td>
<td>558 patients</td>
<td>AUC: 0.967; sensitivity: 87.2%; specificity: 100%</td>
</tr>
<tr>
<td>Watson et al[58]</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of pathologic response of PDAC patients to NAC</td>
<td>LeNet</td>
<td>81 patients (65 for training and validation; 16 for testing)</td>
<td>AUC: 0.785; brier score: 0.174; sensitivity: 81.4%; specificity: 60.4% in test set of hybrid deep learning model</td>
</tr>
<tr>
<td>Muhammad et al[59], 2018</td>
<td>CT</td>
<td>Retrospective</td>
<td>Evaluation of survival hazard of PDAC patients</td>
<td>AlexNet</td>
<td>159 patients</td>
<td>C-index: 0.76; hazard ratio: 9.46 in test set</td>
</tr>
<tr>
<td>Zhang et al[60], 2020</td>
<td>CT</td>
<td>Retrospective</td>
<td>Evaluation of survival probability of PDAC patients</td>
<td>CNN</td>
<td>520 patients</td>
<td>IPA: 11.81%; C-index: 0.651 in testing cohort</td>
</tr>
<tr>
<td>Zhang et al[61], 2020</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of OS of PDAC patients; Evaluation of risk scores to distinguish patients with high or low risk</td>
<td>CNN</td>
<td>98 patients (68 in training cohort; 30 in testing cohort)</td>
<td>AUC: 0.81; hazard ratio: 1.86</td>
</tr>
<tr>
<td>Zhang et al[62], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of 2-yr OS of resectable PDAC patients</td>
<td>CNN</td>
<td>98 patients (68 in training cohort; 30 in testing cohort)</td>
<td>AUC: 0.84; specificity: 68%; sensitivity: 91%</td>
</tr>
<tr>
<td>Yao et al[63], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of survival risk and tumor resection margin of resectable PDAC patients</td>
<td>CNN</td>
<td>205 patients</td>
<td>C-index: 0.705 for survival prediction; balanced accuracy: 73.6%; sensitivity: 81.3%; specificity: 65.9% for resection margin prediction</td>
</tr>
<tr>
<td>Yao et al[63], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of survival risk and tumor resection margin of resectable PDAC patients</td>
<td>CNN</td>
<td>1209 patients</td>
<td>C-index: 0.667 for survival prediction; balanced accuracy: 67.1%; sensitivity: 59.8%; specificity: 74.3% for resection margin prediction</td>
</tr>
<tr>
<td>An et al[64], 2022</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of LNM status and OS in PDAC patients</td>
<td>ResNet-18</td>
<td>148 patients (88 in training cohort, 25 in validation cohort, 35 in testing cohort)</td>
<td>For combined model, AUC: 0.92; accuracy: 86%; sensitivity: 94%; specificity: 78% in testing cohort</td>
</tr>
<tr>
<td>Li et al[65], 2019</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of HMGA2 and C-MYC gene expression status of PDAC patients; Prediction of survival time of patients</td>
<td>CNN</td>
<td>111 patients</td>
<td>Average AUC score: 0.90; accuracy: 95%; sensitivity: 92%; specificity: 98% in C-MYC test with deep features selected by Doctor B; average AUC score: 0.91; accuracy: 88%; sensitivity: 89%; specificity: 88% in HMGA2 test with deep features selected by Doctor B</td>
</tr>
</tbody>
</table>

CT: Computed tomography; US: Ultrasound; PDAC: Pancreatic ductal adenocarcinoma; AIP: Autoimmune pancreatitis; CP: Chronic pancreatitis; NAC: Neoadjuvant chemotherapy; LNM: Lymph node metastasis; OS: Overall survival; AUC: Area under the receiver operating characteristic curve; C-index: Harrell’s concordance index; CNN: Convolutional neural network; IPA: Index of prediction accuracy; DLR: Deep-learning-based radiomics.

DLR model achieved a higher AUC than the HCR model when predicting the response to chemotherapy in CRLM (training: 0.903 vs 0.745; validation: 0.820 vs 0.598). Microsatellite instability (MSI) function is a predictive biomarker for clinical outcomes and predicts responses to adjuvant 5-fluorouracil and immunotherapy in CRC. A DL model that was created using the MobileNetV2 architecture by Zhang et al[69] was adopted to predict the MSI status of CRC based on MR images. With AUC values of 0.868, the best model successfully identified 85.4% of the MSI status, indicating that the suggested model may aid in locating individuals who might benefit from chemotherapy or immunotherapy.

DLR in CRC also emphasizes the need to predict treatment response. For improving NCRT response prediction in locally advanced rectal cancer, Fu et al[70] compared the handcrafted and DL features extracted from pre-treatment diffusion-weighted MR images. The DLR approach produced a mean AUC of 0.73, while the HCR method yielded a mean AUC of 0.64, which demonstrated that DLR may achieve higher classification performance compared with HCRs. To predict the distant metastasis in
locally advanced rectal cancer patients receiving NCRT, Liu et al. [71] exploited the use of a DLR model based on MR images. DLR achieved a C-index of 0.747 and AUC of 0.894 at 3 years. In order to define tumor morphological change for response evaluation in patients with metastatic CRC, Lu et al. [72] offered a DLR study using CNN and recurrent neural network. They discovered that the DL network performed better than the size-based equivalent with C-index (0.649 vs 0.627), and was capable of predicting the early on-treatment response in metastatic CRC. The predictive performance could be improved by the integration of DL network with size-based methodology. LN metastasis, which is a key prognostic factor for CRC, is among the other study topics of CRC. Ding et al. [73] adopted a DLR nomogram based on faster region-based CNN (Faster R-CNN) to predict LN metastasis in patients with CRC. Patient age, Faster R-CNN-detected LN metastasis, and tumor differentiation were predictors in the Faster R-CNN nomogram for predicting LN metastasis, with AUCs in the training and validation sets of 0.862 and 0.920, respectively. Zhao et al. [74] applied a DLR model related with genomics phenotypes for predicting LN metastasis in CRC and showed good performance with AUCs of 0.81, 0.77, and 0.73 in the training, testing and validation sets, respectively. Li et al. [75] examined the performance of the three most popular classification techniques-DL, conventional machine learning, and deep transfer learning-to determine the most efficient way for automatic classification of CRC LN metastases. Deep transfer learning was the most successful, with an accuracy of 0.7583 and AUC of 0.7941. All of these studies have shown that DLR technology has good performance in the prediction and classification of LN metastasis.

The studies exploring the creation of DLR for CRC are summarized in Table 5.

**CHALLENGES AND RECOMMENDATIONS**

In the past several years, with the development of DL technology, the research and application of DLR in tumor diagnosis, treatment and prognosis have been increasing. To perform a systematic evaluation of the status of DLRs for GI cancer, we conducted an extensive review on all original publications between January 1, 2015, and August 30, 2022. Even though several published articles have confirmed the exceptional performance of DLR, there are still many issues that algorithm designers and doctors must address. Below is a list of the challenges and recommendations for DLR in future research summarized by our team.

**Prospective and multi-center studies**

According to the most recent research, most studies on DLR were retrospective and single center. Retrospective studies may have sample selection bias and cannot truly reflect the distribution of clinical cases, which could jeopardize the precision of DLR models. As different centers have different machine parameters, scanning settings, and diagnostic rules, single-center studies limit the generalization of the DLR models. Prospective and multi-center studies can evaluate the reliability and accuracy of the DLR models, enhance their generalization, and bridge the gap between academic studies and clinical applications. Thus, carrying out prospective and multi-center studies is the key to accelerating the clinical application of DLRs models.

**Development of user-friendly DL models**

We found that many physicians do not really want to use DLR methods for related research because the models usually have complex structures, large parameters, poor interpretability, non-existence of gradients, overfitting, and other problems, which limit the promotion and use of DLR technology. Therefore, it is necessary to develop simple and user-friendly models and training schemes for non-professional users. Publication of more source codes and pre-training weights are ways to reduce the development and training difficulty of DL models. For overfitting problems, development of automatic data augmentation schemes and image synthesis schemes can increase the amount of training data. For the black box nature of DL models, attention maps and network dissection schemes can be integrated into the model to improve interpretability.

**Establishment of accessible datasets**

For DLR, the dataset is the new oil. DLR analysis requires a large amount of data to train and validate models; however, most studies are based on private datasets and do not use uniform construction standards, which will hinder the reproducibility of the studies and deployment of DLR models. Thus, a professional data development organization that combines multi-center data should be established. The organization should also standardize the development process of multiple kinds of datasets and make the datasets publicly accessible. Additionally, to reward data contributors, researchers who use these datasets could charge appropriate fees.

**Efficient fusion of multiple features**

DLR is a new technology in the field of AI for medical image analysis. Although its performance is satisfactory, it is not a panacea, especially in the case of extreme shortage of data. Numerous studies
Table 5 Summary of studies using deep-learning-based radiomics for colorectal cancer

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Imaging</th>
<th>Study design</th>
<th>Study aim</th>
<th>DL model</th>
<th>Dataset</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wu et al[67], 2020</td>
<td>CT</td>
<td>Retrospective</td>
<td>Predicting KRAS status in patients with CRC</td>
<td>CNN</td>
<td>Primary cohort: 279 patients; validation cohort: 119 patients</td>
<td>C-index of 0.815 for the primary cohort and 0.832 for the validation cohort</td>
</tr>
<tr>
<td>Wei et al[68], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Predicting the response to chemotherapy in CRLM</td>
<td>ResNet10</td>
<td>192 patients</td>
<td>AUC of DLR: 0.820; AUC of HCR: 0.598</td>
</tr>
<tr>
<td>Zhang et al[69], 2021</td>
<td>MRI</td>
<td>Retrospective</td>
<td>Predicting the MSI status of CRC</td>
<td>MobileNetV2</td>
<td>491 patients</td>
<td>Accuracy: 85.4%; AUC: 0.868</td>
</tr>
<tr>
<td>Fu et al[70], 2021</td>
<td>MRI</td>
<td>Retrospective</td>
<td>Predicting NCRT response in patients with LARC</td>
<td>VGG19</td>
<td>43 patients</td>
<td>AUC of DLR: 0.73; AUC of HCR: 0.64</td>
</tr>
<tr>
<td>Liu et al[71], 2021</td>
<td>MRI</td>
<td>Retrospective</td>
<td>Predicting the distant metastasis of LARC patients receiving NCRT</td>
<td>ResNet18</td>
<td>235 patients</td>
<td>C-index of 0.747 and AUC of 0.894 in the validation cohort</td>
</tr>
<tr>
<td>Lu et al[72], 2021</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of early on-treatment response in mCRC</td>
<td>CNN + RNN</td>
<td>1028 patients</td>
<td>C-index: 0.649</td>
</tr>
<tr>
<td>Ding et al[73], 2020</td>
<td>MRI</td>
<td>Retrospective</td>
<td>Prediction of metastatic LN in CRC</td>
<td>Faster RCNN</td>
<td>545 patients</td>
<td>AUC for training: 0.862; AUC for validation: 0.920</td>
</tr>
<tr>
<td>Zhao et al[74], 2022</td>
<td>CT</td>
<td>Retrospective</td>
<td>Prediction of metastatic LN in CRC</td>
<td>Autoencoder</td>
<td>423 patients</td>
<td>AUC for training: 0.81; AUC for validation: 0.73; AUC for testing: 0.77</td>
</tr>
<tr>
<td>Li et al[75], 2020</td>
<td>MRI</td>
<td>Retrospective</td>
<td>Classification of CRC LN Metastasis images</td>
<td>AlexNet</td>
<td>3364 samples (1646 positive; 1718 negative)</td>
<td>Accuracy: 75.83%; AUC: 0.7941</td>
</tr>
</tbody>
</table>

CRC: Colorectal cancer; CT: Computed tomography; MRI: Magnetic resonance imaging; CRLM: Colorectal liver metastases; MSI: Microsatellite instability; NCRT: Neoadjuvant chemoradiotherapy; LARC: Locally advanced rectal cancer; mCRC: Metastatic colorectal cancer; LN: Lymph node; CNN: Convolutional neural network; RNN: Recurrent neural network; AUC: Area under the receiver operating characteristic curve; C-index: Harrell’s concordance index; DLR: Deep-learning-based radiomics; HCR: Handcrafted radiomics; Faster R-CNN: Faster region-based convolutional neural network.

...have demonstrated that combining HCR and DLR, can result in better performance. Thus, we suggest integrating other clinical features, genomics, handcrafted features, and DL features to build an optimal solution. Moreover, a suitable feature dimensionality reduction scheme should also be adopted to reduce the redundancy of the integrated features. In addition to imaging features, features extracted from clinical data sources, such as gene expression, clinical characteristics, and blood biomarkers, can also be combined to enhance radiomic features.

CONCLUSION

Globally, GI cancers account for a large portion of cancer-related fatalities. For the diagnosis and treatment of GI cancer, DLR can offer a simpler, quicker and more reliable approach. This article is the first comprehensive review on DLR in the GI tract. The status, difficulties, and suggestions discussed in this review can help engineers create optimal radiomics products to support clinical decision-making and offer guidance for diagnosis and treatment of other tumors. Despite the success of DLR in GI cancer, prospective and multi-center studies are still needed. Development of user-friendly DL models, the creation of large public databases, and the fusion of multiple features are also necessary to encourage the clinical application of radiomics.

FOOTNOTES

Author contributions: Wong PK, Chan IN, Yan HM, Gao S, Wong CH, and Yan T collected the literature and wrote the initial manuscript, conceptualized the table and figures, and contributed equally to this work; Yao L, Hu Y, Wang ZR, and Yu HH conceptualized the structure of the text and critically revised the manuscript for important intellectual content; all authors read and approved the final version of the manuscript.

Supported by the Guangdong Basic and Applied Basic Research Fund, Shenzhen Joint Fund (Guangdong-Shenzhen Joint Fund) Guangdong-Hong Kong-Macau Research Team Project, No. 2021B1515130003; Science and Technology
Development Fund of Macau, No. 0026/2022/A; and Project of Xiangyang Science and Technology on Medical and Health Field, No. 2022YL05A.

Conflict-of-interest statement: All authors declare no conflict of interest.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

Country/Territory of origin: China

ORCID number: Pak Kin Wong 0000-0002-7623-6904; In Neng Chan 0000-0002-7015-4510; Hao-Ming Yan 0000-0002-2223-4671; Shao Gao 0000-0002-8232-4134; Chi Hong Wong 0000-0003-2183-4525; Tao Yan 0000-0002-8929-015X; Liang Yao 0000-0003-3377-1509; Ying Hu 0000-0002-3807-3649; Zhong-Ren Wang 0000-0002-3488-7119; Hon Ho Yu 0000-0002-9580-345X.

S-Editor: Chen YL
L-Editor: A
P-Editor: Chen YL

REFERENCES


