

80202_Auto_Edited - check.docx

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

Wong PK *et al.* DPR for gastrointestinal cancer

35

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.

Key Words: Radiomics; Deep learning; Gastrointestinal cancer; Medical imaging

Wong PK, Chan IN, Yan HM, Gao S, Wong CH, Yan T, Yao L, Hu Y, Wang ZR, Yu HH. Deep learning based radiomics for gastrointestinal cancer diagnosis and treatment: A minireview. *World J Gastroenterol* 2022; In press

Core Tip: Radiomics, especially deep-learning-based radiomics (DLR), has revolutionized the diagnosis, assessment and prognosis of gastrointestinal (GI) cancer. This review provides an analysis and status of DLR in GI cancer and identifies future challenges and recommendations.

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 (Figure 1). 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-

² 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.

²⁷ 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 2.

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 2).

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 ¹ 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*^[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*^[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*^[25] designed a 3S-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*^[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*^[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 and lymph node (LN) metastasis. The latter is a significant prognostic factor for ESCC. Wang *et al*^[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*^[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*^[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 LN metastasis, which is an effective prognosis factor of ESCC. Wu *et al*^[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 discriminative 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^[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*^[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*^[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 (SRCC) 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: 0.74, testing: 0.76) and hazard ratios (validation: 3.50, testing: 9.46).

To do a preoperative prediction of early recurrence of patients with advanced GC from CT images, Zhang *et al*^[37] designed a radiomics nomogram that utilized clinical characteristics and radiomics signature containing handcrafted and DL features as input. The radiomics nomogram reached an AUC and accuracy of 0.806 and 0.723, respectively, while having considerable k values of 0.932 for both intra- and inter-reader agreement, exceeding the results obtained by the radiomics signature and clinical modal 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 prognosis 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.

DLRs FOR LIVER CANCER

Primary liver ¹cancer 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 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*^[48] 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

²⁴ Pancreatic cancer is the seventh most deadly cancer worldwide, in which pancreatic adenocarcinoma or ²⁴ 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 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 [\geq 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 UNet 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

¹⁴Colorectal 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.

⁹DLR 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

42 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. The 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 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.

Figure Legends

Figure 1 Estimated global new cases and deaths of major gastrointestinal cancers from 2020 to 2040^[2].

Figure 2 Overview of steps in handcrafted radiomics workflow and steps that can be done with deep learning models. ROI: Region of interest.

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

| Ref. | Imaging | Study design | Study aim | DL model | Dataset | Outcomes |
|--|---------|---------------|--|----------|---|--|
| Takeuchi <i>et al</i> ^[23] , 2021 | CT | Retrospective | Detection of esophageal cancer | VGG16 | 1646 CT images (1500 images for training and validation, 146 for testing) | Accuracy: 84.2%; F value: 74.2%; Sensitivity: 71.1%; Specificity: 90% in test set |
| Hu <i>et al</i> ^[24] , 2021 | CT | Retrospective | Evaluation of response to NCRT to ESCC | ResNet50 | 231 patients (161 in training cohort, 70 in testing cohort) | C-AUC: 0.805; index: 0.805; Accuracy: 77.1%; Sensitivity: 83.9%; Specificity: 71.8% for the testing cohort |
| Ypsilantis <i>et al</i> ^[25] , 2015 | PET | Retrospective | Prediction of response to NAC in patients with esophageal cancer | 3S-CNN | 107 patients | Sensitivity: 80.7%; Specificity: 81.6%; Accuracy: 73.4% |

| | | | | | | |
|---|-----|---------------|--|--------------|---|---|
| Amyar <i>et al</i> ^[26] , 2019 | PET | Retrospective | Prediction of response to radiotherapy in patients with esophageal cancer | 3D RPET-NET | 97 patients | Accuracy: 75.0%; Sensitivity: 76.0%; Specificity: 74.0%; AUC: 0.74 |
| Li <i>et al</i> ^[27] , 2021 | CT | Retrospective | Prediction of treatment response to CCRT among patients with locally advanced TESC | ResNet34 | 306 patients (203 in training cohort, 103 in validation cohort) | AUC: 0.833; PPV: 100% |
| Wang <i>et al</i> ^[28] , 2022 | CT | Retrospective | Prediction of survival rates for patients with esophageal cancer after 3 yr with chemoradiotherapy | DenseNet-169 | 154 patients (116 in training cohort, 38 in validation cohort) | AUC: 0.942; C-index: 0.784 |
| Yang <i>et al</i> ^[29] , 2019 | PET | Retrospective | Identification of esophageal cancer based on | 3D-CNN | 1107 scans | AUC: 0.738 |

| | | | | | | |
|--|------|---------------|---|-------------------------|--------------|----------------|
| Gong <i>et al</i> ^[30] , 2022 | CECT | Retrospective | patients with poor prognosis Prediction of LRFS in esophageal cancer patients after 1 yr of definitive chemoradiotherapy | ResNet18 3D-Densenet | 397 patients | C-index: 0.76 |
| Wu <i>et al</i> ^[31] , 2019 | CT | Retrospective | Prediction of LN status of patients with ESCC | CNN-F | 411 patients | C-index: 0.840 |

18

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.

1 Table 2 Summary of studies using deep-learning-based radiomics for gastric cancer

| Ref. | Imaging | Study design | Study aim | DL model | Dataset | Outcomes |
|---|---------|---------------|--|----------------------|--|---|
| Cui <i>et al</i> ^[32] , 2022 | CT | Retrospective | Prediction of response to NAC in patients with LAGC | DenseNet-121 | 719 patients | C-index: 0.829 |
| Li <i>et al</i> ^[33] , 2022 | CT | Retrospective | Diagnosis and prediction of chemotherapy response to SRCC patients ⁴⁰ | Modified U-Net | 855 patients (598 in training cohort; 257 in testing cohort) | For diagnosis, AUC: 0.786; accuracy: 71.6%; sensitivity: 77.3%; specificity: 69.2% for testing cohort |
| Tan <i>et al</i> ^[34] , 2020 | CT | Retrospective | Prediction of response to chemotherapy in patients with gastric cancer | V-Net | 116 patients | Mean AUC: 0.728 (testing cohort); 0.828 (validation cohort) when using semi-segmentation |
| Hao <i>et al</i> ^[35] , 2021 | CT | Retrospective | Prediction of OS and PFS after gastrectomy; evaluation of effects of variables on survival | Attention-guided VAE | 1061 patients (743 for training; 318 for testing) | C-index of OS: 0.783; C-index of PFS: 0.770 when only using postoperative variables |

| prediction | | | | | |
|---|----|---------------|---|---|---|
| Zhang <i>et al</i> ^[36] , 2021 | CT | Retrospective | Prediction of OS risks of patients with gastric cancer | MMF-FPN 640 patients (337 in training set; 181 in validation set; 122 in test set) | C-index: 0.76; Hazard ratio: 9.46 in test set |
| Zhang <i>et al</i> ^[37] , 2020 | CT | Retrospective | Prediction of early recurrence of patients with AGC | DCNNs 669 patients | ⁴⁹ AUC: 0.806; accuracy: 0.723; sensitivity: 0.827; specificity: 0.667 |
| Guan <i>et al</i> ^[38] , 2022 | CT | Retrospective | Prediction of preoperative status of LNM of gastric cancer patients | ResNet50-RF 347 patients (242 for training; 105 for testing) | AUC: 0.9803; accuracy: 98.10%; sensitivity: 98.39%; specificity: 0.9767% for testing of ResNet50-RF |
| Dong <i>et al</i> ^[39] , 2020 | CT | Retrospective | Prediction of the number of LNM in LAGC | DenseNet-201 730 patients | C-index: 0.822 in validation set |

| | | | | | |
|---|---|--|-------------------------------------|---|---|
| Li <i>et al</i> ^[40] , CT 2020 | Retrospective | Prediction of LNM and prognosis in gastric cancer patients | DCNNs | 204 patients (136 in training set, 68 in test set) | AUC: 0.82 in test set; C-index of OS: 0.67; C-index of PFS: 0.64 |
| Jin <i>et al</i> ^[41] , CT 2021 | Retrospective | Prediction of LNM status in LN stations of gastric cancer patients | ResNet-18 | 1699 patients | Median AUC: 0.876; median Sensitivity: 0.743; median Specificity: 0.936 in validation cohort |
| Sun <i>et al</i> ^[42] , CT 2020 | Stage I: Retrospective; stage II: Validation | Prediction of serosa invasion of AGC patients | DCNNs | 572 patients (252 in training set, 176 in test set I, 144 in test set II) | AUC: 0.87; accuracy: 80%; sensitivity: 0.73; specificity: 0.85 in test set I. AUC: 0.90; accuracy: 85%; sensitivity: 0.75; specificity: 0.93 in test set II |
| Li <i>et al</i> ^[43] , CT 2022 | Retrospective | Evaluation of lymphovascular invasion of localized V3, training, | SqueezeNet, ResNet50, Inception V3, | 1062 patients (728 for 334 accuracy for | AUC: 0.725; sensitivity: 73.2%; specificity: 60.3%; |

| | | | |
|-------------------------|-------------------|--------------|---|
| gastric cancer patients | VGG19, DeepLoc | for testing) | radiomics GRISK model (final model) in testing cohort |
|-------------------------|-------------------|--------------|---|

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.

Table 3 Summary of studies using deep-learning-based radiomics for liver cancer

| Ref. | Imaging | Study design | Study aim | DL model | Dataset | Outcomes |
|-----------------------------------|---------|---------------|-----------------------------------|----------|--|---|
| Ding et al ^[44] , 2021 | CT | Retrospective | Evaluation of HCC differentiation | VGG19 | 1234 patients (799 in training cohort, 248 in validation cohort; 187 in the fused DLRs | AUC: 0.8042; accuracy: 72.73%; sensitivity: 70.75%; specificity: 75.31% in testing cohort |

| | | | | independent testing cohort | model |
|--|----|---------------|---|--|--|
| Peng <i>et al</i> ^[45] , 2020 | CT | Retrospective | Prediction of different treatment responses to TACE in HCC patients | 789 patients (562 in training cohort; 89 in validation cohort 1; 138 in validation cohort 2) | Accuracy: 85.1% in validation cohort 1; 82.8% in validation cohort 2 |
| Peng <i>et al</i> ^[46] , 2021 | CT | Retrospective | Prediction of initial response to TACE in HCC patients | 310 patients (139 in training cohort; 171 in validation cohort) | AUC: 0.994 |
| Wei <i>et al</i> ^[47] , 2021 | CT | Retrospective | Prediction of OS of HCC patients treated with SBRT | 167 patients | C-index: 0.650 in cross validation |
| Liu <i>et al</i> ^[48] , 2020 | US | Retrospective | Prediction of PFS of HCC patients | 214 RFA patients (149 for training; 65 for validation) | C-index of RFA: 0.726; C-index of surgical |

| | | | |
|---|--|---|--|
| | treated with RFA or resection | 65 validation), 205 SR patients (144 for training; 61 for validation) | for resection: 0.726 |
| Wang <i>et al</i> ^[49] , 2019 | Retrospective Prediction of early recurrence of HCC patients | ResNet 167 patients | AUC of best model: 0.825 |
| Wang <i>et al</i> ^[50] , 2020 | Retrospective Prediction of early recurrence of HCC patients | ResNet 167 patients | For the best model with joint loss function, AUC: 0.8331; accuracy: 80.49% |
| He <i>et al</i> ^[51] , MRI 2021 | and Retrospective pathological data Evaluation of HCC recurrence risk of liver transplantation recipients | U-net, CapsNet 109 patients (87 for training; 22 for testing) | Total accuracy: 82%; recall: 80%; precision: 89%; AUC: 0.87; F-1 score: 84% |
| Jiang <i>et al</i> ^[52] , 2021 | Retrospective Prediction of microvascular | 3D-CNN 405 patients (324 in training set, 81 | AUC: 0.906; sensitivity: 75.7%; specificity: 93.2%; |

| | | | | | | |
|--|---------------|---|------------------|---|--|-----|
| | | invasion status of HCC patients | | in validation set) | accuracy: score: 87.2% in validation set | F-1 |
| Wang <i>et al</i> ^[53] , 2022 | Retrospective | Prediction of microvascular invasion status of HCC patients | Transformer, CNN | 138 patients | For arterial phase images in validation set, AUC: 0.9223; Average accuracy: 86.78% | |
| Fu <i>et al</i> ^[54] , 2021 | Retrospective | Prediction of macrovascular invasion status in HCC patients | Modified Net | U- 366 patients (281 in training cohort, 85 in validation cohort) | AUC: 0.836 in validation | |

44 CT: Computed tomography; MRI: Magnetic resonance imaging; US: Ultrasound; HCC: Hepatocellular carcinoma; TACE: 28 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.

1 Table 4 Summary of studies using deep-learning-based radiomics for pancreatic cancer

| Ref. | Imaging | Study design | Study aim | DL model | Dataset | Outcomes |
|--|---------|---------------|---|-----------|--|---|
| Ziegelmayr <i>et al</i> ^[55] , 2020 | CT | Retrospective | Identification of PDAC and AIP | VGG19 | 86 patients (44 AIP patients and 42 PDAC patients) | Sensitivity: 89%; specificity: 83%; AUC: 0.90 |
| Liao <i>et al</i> ^[56] , 2022 | CT | Retrospective | Identification of PDAC, non-cancerous pancreatic diseases and normal pancreas | CNN | 3120 images (1872 for training, 624 for validation, 624 for testing) | Sensitivity: 89.9%; specificity: 91.3%; AUC: 0.960 when distinguishing between PDAC and control group |
| Tong <i>et al</i> ^[57] , 2022 | US | Retrospective | Identification of PDAC and CP | ResNet-50 | 558 patients | AUC: 0.967; sensitivity: 87.2%; specificity: 100% |
| Watson <i>et al</i> ^[58] | CT | Retrospective | Prediction of pathologic response of | LeNet | 81 patients (65 for training and validation; 16 for test set of hybrid | AUC: 0.785; brier score: 0.174; sensitivity: 81.4%; specificity: 60.4% in test set of hybrid |

| | | | | | | |
|--|----|---------------|---|---------|---|---|
| | | | PDAC patients to NAC | | testing) | deep learning model |
| Muhammad <i>et al</i> ^[59] , 2018 | CT | Retrospective | Evaluation of survival hazard of PDAC patients | AlexNet | 159 patients | C-index: 0.76; Hazard ratio: 9.46 in test set |
| Zhang <i>et al</i> ^[60] , 2020 | CT | Retrospective | Evaluation of survival probability of PDAC patients | CNN | 520 patients | IPA: 11.81%, C-index: 0.651 in testing cohort |
| Zhang <i>et al</i> ^[61] , 2020 | CT | Retrospective | Prediction of OS of PDAC patients; Evaluation of risk scores to distinguish patients with high or low | CNN | 98 patients (68 in training cohort; 30 in testing cohort) | AUC: 0.81; Hazard Ratio: 1.86 |

| | | | | | |
|-------------------------------------|---------------|--|-----|---|--|
| Zhang <i>et al</i> [62], CT 2021 | Retrospective | Prediction of 2-yr OS of resectable PDAC patients | CNN | 98 patients (68 in training cohort; 30 in testing cohort) | AUC: 0.84; specificity: 68%; sensitivity: 91%. |
| Yao <i>et al</i> [63], CT 2021 | Retrospective | Prediction of survival risk and tumor resection margin resectable PDAC patients | CNN | 205 patients | C-index: 0.705 for survival prediction; balanced accuracy: 73.6%, sensitivity: 81.3%, specificity: 65.9% for resection margin prediction |
| Yao <i>et al</i> [63], CT 2021 | Retrospective | Prediction of survival risk and tumor resection margin of resectable PDAC patients | CNN | 1209 patients | C-index: 0.667 for survival prediction; balanced accuracy: 67.1%; sensitivity: 59.8%; specificity: 74.3% for resection margin prediction |

| | | | |
|---------------------------------------|---------------|---|--|
| An et al ^[64] , CT 2022 | Retrospective | PDAC patients | |
| | | Prediction of LNM and OS in PDAC patients | 148 patients (88 in training cohort, 25 in validation cohort, 35 in testing cohort) |
| Li et al ^[65] , 2019 CT | Retrospective | Prediction of CNN | 111 patients |
| | | HMGGA2 and C-MYC gene expression | Average AUC score: 0.90; accuracy: 95%; sensitivity: 92%; specificity: 98% in C-MYC test with deep features |
| | | status of PDAC patients; | selected by Doctor B; average AUC score: 0.91; accuracy: 88%; sensitivity: 89%; specificity: 88% in HMGGA2 test with deep features |
| | | Prediction of survival time of patients | selected by Doctor B |

37

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

20

receiver operating characteristic curve; C-index: Harrell's concordance index; CNN: Convolutional neural network; IPA: Index of prediction accuracy; DLR: Deep-learning-based radiomics.

1 Table 5 Summary of studies using deep-learning-based radiomics for colorectal cancer

| Ref. | Imaging | Study design | Study aim | DL model | Dataset | Outcomes |
|---|---------|---------------|---|-------------|---|---|
| Wu <i>et al</i> ^[67] , 2020 | CT | Retrospective | Predicting KRAS status in patients with CRC. | CNN | Primary cohort: 279 patients; Validation cohort: 119 patients | C-index of 0.815 for the primary cohort and 0.832 for the validation cohort |
| Wei <i>et al</i> ^[68] , 2021 | CT | Retrospective | Predicting the response to chemotherapy in CRLM | ResNet10 | 192 patients | AUC of DLR: 0.820; AUC of HCR: 0.598 |
| Zhang <i>et al</i> ^[69] , 2021 | MRI | Retrospective | Predicting the MSI status of | MobileNetV2 | 491 patients | Accuracy: 85.4%; AUC: 0.868 |

| | | | | | | | |
|--|-----|---------------|---|-------------|---------------|---|--|
| Fu <i>et al</i> ^[70] , 2020 | MRI | Retrospective | CRC | | | | |
| | | | Predicting NCRT response in patients with LARC | VGG19 | 43 patients | AUC of DLR: 0.73; AUC of HCR: 0.64 | |
| Liu <i>et al</i> ^[71] , 2021 | MRI | Retrospective | Predicting the distant metastasis of LARC patients receiving NCRT | ResNet18 | 235 patients | ⁴ C-index of 0.747 and AUC of 0.894 in the validation cohort | |
| | | | Prediction of early on- treatment response in mCRC | CNN + RNN | 1028 patients | C-index: 0.649 | |
| Ding <i>et al</i> ^[73] , 2020 | MRI | Retrospective | Prediction of metastatic LN in CRC | Faster RCNN | 545 patients | AUC for training: 0.862; AUC for validation: 0.920 | |

| | | | | | | |
|--|-----|---------------|---|-------------|---|---|
| Zhao <i>et al</i> ^[74] , 2022 | CT | Retrospective | Prediction of metastatic LN in CRC | Autoencoder | 423 patients | AUC for training: 0.81; AUC for validation: 0.73; AUC for testing: 0.77 |
| Li <i>et al</i> ^[75] , 2020 | MRI | Retrospective | Classification of CRC Metastasis images | AlexNet LN | 3364 samples (1646 positive; 1718 negative) | Accuracy: 75.83%; AUC: 0.7941 |

³³ CRC: Colorectal cancer; CT: Computed tomography; MRI: Magnetic resonance imaging; CRLM: Colorectal liver metastases; MSI: Microsatellite instability; ¹⁶ NCRT: Neoadjuvant chemoradiotherapy; LARC: Locally advanced rectal cancer; mCRC: Metastatic colorectal cancer; LN: Lymph node; CNN: Convolutional neural network; RNN: Recurrent neural network; ³ 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.

12%

SIMILARITY INDEX

PRIMARY SOURCES

- 1

www.ncbi.nlm.nih.gov
Internet

326 words — 3%
- 2

Parnian Afshar, Arash Mohammadi, Konstantinos N. Plataniotis, Anastasia Oikonomou, Habib Benali. "From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities", IEEE Signal Processing Magazine, 2019
Crossref

49 words — 1%
- 3

www.gutnliver.org
Internet

38 words — < 1%
- 4

Hamid Abdollahi, Erika Chin, Haley Dean Clark, Derek E Hyde, Steven Thomas, Jonn Wu, Carlos Felipe Uribe, Arman Rahmim. "Radiomics-guided radiation therapy: opportunities and challenges", Physics in Medicine & Biology, 2022
Crossref

33 words — < 1%
- 5

Tiancheng He, Joy Nolte Fong, Linda W. Moore, Chika F. Ezeana et al. "An Imageomics and Multi-Network based Deep Learning Model for Risk Assessment of Liver Transplantation for Hepatocellular Cancer", Computerized Medical Imaging and Graphics, 2021
Crossref

32 words — < 1%

6 Jiaojiao Zhao, Han Wang, Yin Zhang, Rui Wang et al. "Deep learning radiomics model related with genomics phenotypes for lymph node metastasis prediction in colorectal cancer", Radiotherapy and Oncology, 2021 28 words — < 1%

Crossref

7 "ECR 2018 - BOOK OF ABSTRACTS", Insights into Imaging, 2018 27 words — < 1%

Crossref

8 "Program Abstracts from the GSA 2019 Annual Scientific Meeting "Strength in Age—Harnessing the Power of Networks"", Innovation in Aging, 2019 25 words — < 1%

Crossref

9 Jie Fu, Xinran Zhong, Ning Li, Ritchell Van Dams, John Lewis, Kyunghyun Sung, Ann C Raldow, Jing Jin, X Sharon Qi. "Deep learning-based radiomic features for improving neoadjuvant chemoradiation response prediction in locally advanced rectal cancer", Physics in Medicine & Biology, 2020 24 words — < 1%

Crossref

10 f6publishing.blob.core.windows.net 23 words — < 1%

Internet

11 "Scientific Abstracts and Sessions", Medical Physics, 2017 22 words — < 1%

Crossref

12 Liwen Zhang, Di Dong, Lianzhen Zhong, Cong Li, Chaoen Hu, Xin Yang, Zaiyi Liu, Rongpin Wang, Junlin Zhou, Jie Tian. "Multi-focus Network to Decode Imaging Phenotype for Overall Survival Prediction of Gastric Cancer Patients.", IEEE Journal of Biomedical and Health Informatics, 2021 22 words — < 1%

Crossref

13 Jie Gong, Wencheng Zhang, Wei Huang, Ye Liao, Yutian Yin, Mei Shi, Wei Qin, Lina Zhao. "CT-based radiomics nomogram may predict local recurrence-free survival in esophageal cancer patients receiving definitive chemoradiation or radiotherapy: a multicenter study", Radiotherapy and Oncology, 2022
Crossref 21 words — < 1%

14 www.nature.com
Internet 19 words — < 1%

15 Fei Liu, Dan Liu, Kun Wang, Xiaohua Xie et al. "Deep Learning Radiomics Based on Contrast-Enhanced Ultrasound Might Optimize Curative Treatments for Very-Early or Early-Stage Hepatocellular Carcinoma Patients", Liver Cancer, 2020
Crossref 18 words — < 1%

16 www.researchsquare.com
Internet 18 words — < 1%

17 www.science.gov
Internet 17 words — < 1%

18 www.wjgnet.com
Internet 16 words — < 1%

19 link.springer.com
Internet 15 words — < 1%

20 www.thno.org
Internet 15 words — < 1%

21 Qiong Li, Qiu-Xia Feng, Liang Qi, Chang Liu, Jing Zhang, Guang Yang, Yu-Dong Zhang, Xi-Sheng Liu. "Prognostic aspects of lymphovascular invasion in localized gastric cancer: new insights into the radiomics and deep

transfer learning from contrast-enhanced CT imaging",
Abdominal Radiology, 2021

Crossref

22 Xiao Guan, Na Lu, Jianping Zhang. "Computed Tomography-Based Deep Learning Nomogram Can Accurately Predict Lymph Node Metastasis in Gastric Cancer", Digestive Diseases and Sciences, 2022

14 words — < 1%

Crossref

23 www.e-sciencecentral.org

Internet

14 words — < 1%

24 Jiawen Yao, Yu Shi, Kai Cao, Le Lu, Jianping Lu, Qike Song, Gang Jin, Jing Xiao, Yang Hou, Ling Zhang. "DeepPrognosis: Preoperative Prediction of Pancreatic Cancer Survival and Surgical Margin via Comprehensive Understanding of Dynamic Contrast-Enhanced CT Imaging and Tumor-Vascular Contact Parsing", Medical Image Analysis, 2021

13 words — < 1%

Crossref

25 Anne-Sophie Mehdorn, Timo Gemoll, Hauke Busch, Katharina Kern et al. "Biomarkers in Liquid Biopsies for Prediction of Early Liver Metastases in Pancreatic Cancer", Cancers, 2022

12 words — < 1%

Crossref

26 Liangjian Lu, Albert Koulman, Clive J. Petry, Benjamin Jenkins et al. "An Unbiased Lipidomics Approach Identifies Early Second Trimester Lipids Predictive of Maternal Glycemic Traits and Gestational Diabetes Mellitus", Diabetes Care, 2016

12 words — < 1%

Crossref

27 www.minervamedica.it

Internet

12 words — < 1%

28 Luwen Mu, Tao Pan, Ning Lyu, Lin Sun, Shaolong Li, Qiankun Xie, Haijing Deng, Peihong Wu, Hui Liu, Ming Zhao. "CT-guided percutaneous radiofrequency ablation for lung neoplasms adjacent to the pericardium", Lung Cancer, 2018

Crossref

11 words — < 1%

29 cps.org.tw

Internet

11 words — < 1%

30 "Medical Image Computing and Computer Assisted Intervention – MICCAI 2020", Springer Science and Business Media LLC, 2020

Crossref

10 words — < 1%

31 Dong Tian, Hao-Ji Yan, Haruhiko Shiiya, Masaaki Sato, Aya Shinozaki-Ushiku, Jun Nakajima. "Machine learning-based radiomic computed tomography phenotyping of thymic epithelial tumors: Predicting pathological and survival outcomes", The Journal of Thoracic and Cardiovascular Surgery, 2022

Crossref

10 words — < 1%

32 bmcpulmmed.biomedcentral.com

Internet

10 words — < 1%

33 bmcsurg.biomedcentral.com

Internet

10 words — < 1%

34 coek.info

Internet

10 words — < 1%

35 core.ac.uk

Internet

10 words — < 1%

36 www.mdpi.com

Internet

10 words — < 1%

38 Hongwei Zhang, Jiaqi Lu, Yingying Lu, Jiayi Zhou, Zehua Wang, Haiou Liu, Congjian Xu. "Prognostic significance and predictors of the system inflammation score in ovarian clear cell carcinoma", PLOS ONE, 2017
Crossref

39 Jing-wen Tan, Lan Wang, Yong Chen, WenQi Xi, Jun Ji, Lingyun Wang, Xin Xu, Long-kuan Zou, Jian-xing Feng, Jun Zhang, Huan Zhang. "Predicting Chemotherapeutic Response for Far-advanced Gastric Cancer by Radiomics with Deep Learning Semi-automatic Segmentation", Journal of Cancer, 2020
Crossref

40 Kaustav Bera, Nathaniel Braman, Amit Gupta, Vamsidhar Velcheti, Anant Madabhushi. "Predicting cancer outcomes with radiomics and artificial intelligence in radiology", Nature Reviews Clinical Oncology, 2021
Crossref

41 Lei Ding, Guangwei Liu, Xianxiang Zhang, Shanglong Liu, Shuai Li, Zhengdong Zhang, Yuting Guo, Yun Lu. "A deep learning nomogram kit for predicting metastatic lymph nodes in rectal cancer", Cancer Medicine, 2020
Crossref

42 Xiaomei Wu, Yajun Li, Xin Chen, Yanqi Huang et al. "Deep Learning Features Improve the Performance of a Radiomics Signature for Predicting KRAS Status in Patients with Colorectal Cancer", Academic Radiology, 2020
Crossref

-
- 43 academic.oup.com 9 words — < 1%
Internet
-
- 44 office.wjgnet.com 9 words — < 1%
Internet
-
- 45 ovarianresearch.biomedcentral.com 9 words — < 1%
Internet
-
- 46 123dok.co 8 words — < 1%
Internet
-
- 47 Hejiang Cai, Suning Liu, Haiyun Shi, Zhaoqiang Zhou, Shijie Jiang, Vladan Babovic. "Toward improved lumped groundwater level predictions at catchment scale: mutual integration of water balance mechanism and deep learning method", Journal of Hydrology, 2022 8 words — < 1%
Crossref
-
- 48 Shahrokh F. Shariat, Michael W. Kattan. "Chapter 10 Nomograms for Prostate Cancer", Springer Science and Business Media LLC, 2009 8 words — < 1%
Crossref
-
- 49 Taylor A. Burke, Brooke A. Ammerman, Ross Jacobucci. "The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review", Journal of Affective Disorders, 2018 8 words — < 1%
Crossref
-
- 50 Yihuai Hu, Chenyi Xie, Hong Yang, Joshua W.K. Ho et al. "Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma", Radiotherapy and Oncology, 2021 8 words — < 1%
Crossref
-

| | | |
|----|--|----------------|
| 51 | medinform.jmir.org Internet | 8 words — < 1% |
| 52 | synapse.koreamed.org Internet | 8 words — < 1% |
| 53 | usir.salford.ac.uk Internet | 8 words — < 1% |
| 54 | worldwidescience.org Internet | 8 words — < 1% |
| 55 | www.dovepress.com Internet | 8 words — < 1% |
| 56 | www.spandidos-publications.com Internet | 8 words — < 1% |

EXCLUDE QUOTES OFF
EXCLUDE BIBLIOGRAPHY OFF

EXCLUDE SOURCES OFF
EXCLUDE MATCHES OFF