Abstract

Cancers are the main cause of death worldwide. Assessment of tumors by imaging is key to formulating treatment plans. More noninvasive markers of imaging are needed to evaluate primary tumors and treatment responses to achieve individualized diagnosis and treatment. However, conventional image assessment has limitations based on differences in individual radiologists’ interpretation and inability to read hidden high-dimensional features. Artificial intelligence (AI) can provide quantitative, rather than qualitative, assessments based on radiographic characteristics, and has been frequently and successfully applied in the medical image analysis field. In this review, we summarize research progress in the use of AI for evaluating the diagnosis and treatment response of patients with common tumors, including gliomas, lung cancer, liver cancer and colorectal cancer. We further discuss several issues that must be solved in the future to make better use of AI in tumor diagnosis and assessment of treatment response.

Keywords: artificial intelligence, cancer imaging, diagnosis, treatment response

1. INTRODUCTION

Cancer is a major public health problem and is the main cause of death worldwide; gliomas, lung cancer, liver cancer and colorectal cancer (CRC) are the most common tumors endangering human health [1]. Surgery, radiotherapy, chemotherapy and immunotherapy are recommended anticancer therapies based on precise diagnosis from medical images. However, conventional image assessment has recognized limitations, owing to differences in individual radiologists’ interpretation and inability to read hidden high-dimensional features. Artificial intelligence (AI) can recognize complex patterns in imaging automatically and provide quantitative, rather than qualitative, assessment of radiographic characteristics [2]. AI has been frequently and successfully applied in the medical image analysis field [3, 4].

AI is a new technical science that studies and develops theories, methods, technologies and application systems to simulate, extend and expand human intelligence. Machine learning is a branch of AI that aims to automatically extract trends and features from data and help make wise decisions. According to whether labels are needed, machine learning algorithms can be divided into supervised learning, semisupervised learning and unsupervised learning. In medical imaging, the use of radiomics or deep learning for differential diagnosis, segmentation and detection of tumors is a promising field.

Radiomics [5] is a research field aimed at establishing models that may play important roles in diagnosis, prognosis and prediction. First-order and high-order image features of medical images are key in helping the models analyze and make correct judgements. Because the tumor target is relatively small, and the internal heterogeneity of images is high, researchers often use radiomics for differential diagnosis of tumors [6]. The main steps in radiomics include data standardization, region of interest (ROI) placement, feature extraction [7], feature standardization, feature selection and model construction. With the development of radiomics, several available tools [8, 9] help researchers better perform radiomics.

However, because the neural networks have more parameters than traditional machine learning models, the performance is often better on larger datasets.
Deep learning is a nonlinear parameter model, and most deep learning algorithms are based on neural networks. Convolutional neural networks (CNNs) are the main neural networks used in computer vision and medical imaging. Owing to more nonlinear combinations and hyperparameters, deep learning in large datasets tends to achieve good performance [10]. Because labeling medical image data is difficult, deep learning strategies based on semisupervised or self-supervised methods are gradually being applied [11, 12]. Although semisupervised or self-supervised methods are not as effective as supervised learning, these two methods have great potential for future applications, because of the difficulty of medical data annotation. The transformer framework is another branch of deep learning, which was proposed by Google in 2017. On the basis of the transformer framework, Vision transformer (VIT) has been proposed [13]. VIT is composed mainly of multihead attention, position encoding, feed forward, input/output embedding and normalization layers. Compared with CNN, VIT has advantages in some tasks, including differential diagnoses of diseases, lesion segmentation and image registration [14]. Although deep learning performs better in large datasets, the deep learning algorithm is akin to a “black box” and cannot currently be fully explained from the mathematical theory level. Therefore, the interpretability of the characteristics of deep learning is poor. Although the locations focused on by the network (such as lesion locations and regions) can be revealed through feature map visualization, the meaning of deep learning features cannot be qualitatively explained. Even the focus of the network reflected by the feature maps is not the actual focus of physicians. Because radiomics and deep learning each have their own advantages in feature dimension and interpretability, the method of joint feature modeling is expected to provide improvements in solving existing medical problems [15].

Therefore, this article provides a summary of current clinical applications of AI in determining the diagnosis and treatment response for several cancers (Figure 1).

2. GLIOMAS

Gliomas, the most common primary intracranial tumors, represent approximately 81% of primary malignant intracranial tumors [16, 17]. Compared with other cancers, gliomas require a relatively high resource burden for diagnosis and management [18]. For the detection, diagnosis, classification and prognosis prediction of gliomas, magnetic resonance imaging (MRI) is the imaging method of choice [19]. Advanced MRI provides more details of gliomas than conventional MRI. However, the assessment of multiparameter MRI is time consuming and challenging for radiologists. With the development of computer technology, AI reveals more invisible details and is useful in the diagnosis and management of gliomas.

2.1 Diagnosis of gliomas

Automatic glioma detection and categorization are key components of preoperative evaluation. Biopsy is the gold standard for diagnosis, but the results across studies are widely heterogeneous [20]. Preoperative MRI

![Figure 1](https://example.com/figure1.png)
plays an important role in the detection and categorization of glioma [19]. Glioma and brain metastasis are the two most common malignant intracranial tumors in adults, and differentiating the two is challenging for radiologists [21]. The application of AI techniques helps radiologists improve the discriminatory performance, with an AUC >0.80 [22, 23]. The differential diagnosis of glioma, lymphoma, meningioma and pituitary tumors in adults and children has also been improved by AI techniques [24-26].

On the basis of histological appearance, gliomas have been assigned WHO grades I, II, III and IV, which indicate different degrees of malignancy [27]. Limited by the diversity of MRI manifestations, radiologists have limited ability to distinguish high-grade from low-grade gliomas. Radiomics, machine learning or deep learning models have demonstrated good performance in glioma grading [28-30]. Application of novel deep neural network models to extract the common and distinctive information from multimodal data has effectively combined multimodal data information and increased the accuracy of glioma grading [31]. Although deep learning models have good or even excellent performance, they are limited by the large number of MRI parameters and complex calculations; the problem of low speed must be solved. A lightweight 3D UNet deep learning framework has been developed to solve the dilemma of low speed by simple, effective and noninvasive diagnostic approaches [32].

2.2 Genetics and molecular marker detection

Molecular biomarkers in gliomas have shown more potential than WHO grading in predicting prognosis and accurately classifying gliomas; examples include IDH mutation, 1p/19q codeletion and O-6-methylguanine-DNA methyltransferase [27, 33]. The new classification based on molecular biomarkers was added to the World Health Organization criteria [34]. Accurate classification of molecular subtypes of gliomas is crucial for treatment planning and prognostic assessment, and tumor biopsy is the gold standard. However, tumor biopsy is invasive and not feasible in all patients. Noninvasive recognition of molecular subtypes is desirable in routine clinical practice.

Indeed, even experienced radiologists find that accurately classifying molecular subtypes on the basis of preoperative MR images is challenging [35, 36]. Machine learning or deep learning models have shown good performance in molecular biomarker prediction, with an accuracy of 83.8%–97.14 [37-39]. However, the diagnostic accuracy of AI algorithms based on routine preoperative MRI sequences still must be further improved before they can be reliably applied in routine clinical practice [35, 40]. The inclusion of advanced MR techniques has improved the diagnostic performance of AI models. Bumes et al. have applied magnetic resonance spectroscopy to build a machine learning model and achieved good diagnostic performance (sensitivity: 82.6%, specificity: 72.7%) [41]. Several advanced MRI techniques have also been applied to build AI models, such as DWI, ADC, T1 perfusion and ASL.

Therefore, the diagnosis and classification of gliomas are challenging for radiologists, and the development of AI techniques and advanced MRI sequences have shown tremendous improvement and promising results in clinical applications.

2.3 Monitoring response to treatment

Predicting the survival of patients with glioma has substantial clinical value and poses challenges for radiologists in routine clinical practice. Radiomics models based on multimodal images perform better than monomodal images [42]. Advanced MRI images (diffusion, perfusion, arterial spin labeling and MR spectroscopy) have also been used for prognostication [43-46]. However, missing MRI sequences in the multimodal imaging feature-based prediction model limit its application and generalization in routine clinical use. Deep learning models can synthesize missing sequences to ensure the successful implementation of models [47-49]. The post-treatment imaging feature-based prediction model is generalizable and performs better than the pretreatment imaging feature-based prediction model [50]. Multimodal and multi-time-point glioma segmentation is time consuming, and its clinical application is immensely challenging. The deep learning model performs rapidly and accurately in the segmentation of both preoperative and postoperative gliomas [51]. Li et al. have developed a deep learning model for predicting the prognosis of gliomas from whole-brain MRI without tumor segmentation and achieved good performance [52].

Distinguishing pseudoprogression from true tumor progression is challenging and has critical implications for treatment protocols [53]. MRI with gadolinium contrast is the standard imaging method for determining tumor growth. Because traditional MRI cannot reliably distinguish pseudoprogression from true tumor progression, several approaches, including advanced MRI imaging, radiomics and machine learning, have been applied to solve this problem. AI models with all available MRI parameters have also demonstrated better diagnostic performance in distinguishing pseudoprogression from true tumor progression than a model with a single MRI parameter [54, 55].

A summary of the AI studies in gliomas can be found in Table 1.

3. LUNG CANCER

Currently, lung cancer is the second most common cancer and the leading cause of worldwide cancer-associated death [1]. Non-small cell lung cancer (NSCLC) makes up 85–90% of the different pathological types of lung cancer, including adenocarcinoma, squamous cell carcinoma, large cell carcinoma and squamous adenocarcinoma [56]. Early diagnosis, screening and personalized
<table>
<thead>
<tr>
<th>Reference</th>
<th>Application</th>
<th>No. of cases</th>
<th>Imaging modality</th>
<th>Algorithm</th>
<th>Image feature type</th>
<th>Type of validation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conte et al. [47]</td>
<td>Synthesizing missing MRI sequences</td>
<td>210</td>
<td>T1 WI, T2 WI, FLAIR, T1-CE</td>
<td>Generative adversarial networks</td>
<td>Deep learning</td>
<td>Internal validation</td>
<td>MSE of 0.004–0.103</td>
</tr>
<tr>
<td>Bathla et al. [26]</td>
<td>Differentiating between glioblastoma and lymphoma</td>
<td>94</td>
<td>T1 WI, T2 WI, DWI</td>
<td>Machine learning</td>
<td>Radiomics</td>
<td>Fivefold cross-validation</td>
<td>AUC of 0.977</td>
</tr>
<tr>
<td>Ahammed et al. [30]</td>
<td>Glioma grade identification</td>
<td>557</td>
<td>T2 WI</td>
<td>Wndchrm tool based classifier and VGG-19 DNN</td>
<td>Wndchrm feature and deep learning</td>
<td>Internal validation</td>
<td>Accuracy of 92.86% for the Wndchrm classifier and 98.25% for VGG-19 DNN classifier</td>
</tr>
<tr>
<td>Bangalore et al. [37]</td>
<td>Classification of IDH mutation status</td>
<td>214</td>
<td>T1 WI, T2 WI, FLAIR, T1-CE</td>
<td>Deep-learning network</td>
<td>Deep learning</td>
<td>Threefold cross-validation</td>
<td>Sensitivity of 0.98±0.02, specificity of 0.97±0.001, and AUC of 0.99±0.01</td>
</tr>
<tr>
<td>Bumes et al. [41]</td>
<td>Classification of IDH mutation status</td>
<td>101</td>
<td>MRS</td>
<td>Machine learning</td>
<td>Voxel for MRS</td>
<td>External validation</td>
<td>Sensitivity of 82.6% and specificity of 72.7%</td>
</tr>
<tr>
<td>Cheng et al. [31]</td>
<td>Glioma grade identification</td>
<td>350</td>
<td>T1 W, T2 WI, FLAIR, T1-CE</td>
<td>MMD-VAE model</td>
<td>Radiomics</td>
<td>External validation</td>
<td>AUC of 0.9611–0.9939</td>
</tr>
<tr>
<td>Kim et al. [55]</td>
<td>Differentiation of pseudoprogression from early tumor progression</td>
<td>118</td>
<td>T1-CE, FLAIR, ADC and CBV</td>
<td>Radiomics model</td>
<td>Radiomics</td>
<td>External validation</td>
<td>AUC of 0.85</td>
</tr>
<tr>
<td>Li et al. [52]</td>
<td>Prediction of glioma overall survival</td>
<td>1556</td>
<td>T1 WI, T2 WI, FLAIR, T1-CE</td>
<td>DeepRisk model</td>
<td>Whole-brain MRI without tumor segmentation</td>
<td>External validation</td>
<td>AUC of 0.77–0.94</td>
</tr>
<tr>
<td>George et al. [50]</td>
<td>Prediction of progression-free survival and overall survival</td>
<td>113</td>
<td>Pretreatment and first on-treatment time point MR imaging</td>
<td>Random survival forest algorithm</td>
<td>Radiomics</td>
<td>External validation</td>
<td>Concordance index of 0.680–0.715</td>
</tr>
<tr>
<td>Yu et al. [32]</td>
<td>Glioma grade identification</td>
<td>560</td>
<td>T1 WI, T2 WI, FLAIR, T1-CE</td>
<td>Lightweight 3D UNet deep learning framework</td>
<td>Deep learning</td>
<td>Internal validation</td>
<td>Accuracy of 89.29%</td>
</tr>
</tbody>
</table>

MSE, mean squared error; DNN, deep convolutional neural network; IDH, isocitrate dehydrogenase; T1-CE, contrast-enhanced T1 WI; MRS, magnetic resonance spectroscopy; MMD-VAE, multimodal disentangled variational autoencoder.
3.1 Diagnosis of NSCLC
Screening is becoming an important part of early lung cancer diagnosis. In 2020, Maldonado et al. [58] constructed a radiomics model with an AUC of 0.90 to predict the risk of incidentally discovered indeterminate pulmonary nodules. The model performs well and may facilitate the management of indeterminate pulmonary nodules. Meng et al. [59] have established a nomogram to identify indolent and invasive lung adenocarcinomas manifesting as GGNs by combining radiomics, CT semantic features and clinical information, with an AUC reaching 0.946 in the test set. This radiomics nomogram can accurately predict the invasiveness of GGNs before surgery to assist clinicians in formulating individualized treatment strategies. Hu et al. [60] have differentiated benign from malignant GGNs by using a fusion model combined with deep learning and radiomics features, with an accuracy of 75.6%, a value higher than those of the DNN model and radiomics model individually. In addition, the authors have found that it was helpful to build DNN model with limited dataset by using transfer learning.

Different histopathological subtypes of lung cancer determine the choice of treatment modality. Pathological diagnosis is informative but is also invasive. In addition, missampling due to tumor heterogeneity can affect biopsy results. These limitations have encouraged researchers to develop noninvasive and accurate imaging markers to assess the overall tumor and to complement the deficiencies in invasive pathological examinations. Studies have attempted to predict histopathological subtypes of lung cancer on the basis of radiomics. Zhu et al. [61] have extracted five imaging features to classify adenocarcinoma vs. squamous cell carcinoma in NSCLC, and achieved an AUC of 0.893 in the test set, thus demonstrating the potential of noninvasive imaging in the histopathological classification of lung cancer. Maldonado et al. [62] have developed a noninvasive computer-aided nodule risk assessment tool to stratify the adenocarcinoma spectrum. Song et al. [63] have interrogated an entire tumor in a non-invasive manner by establishing a radiomics model, which has provided additional diagnostic information to recognize the micropapillary component and thus guide treatment planning.

3.2 Genetics and molecular marker detection
Currently, targeted therapy and immune therapy for lung cancer have been widely used in clinical settings, and the selection of patients who would benefit depends mainly on genetic mutation status (such as EGFR or ALK) and the expression of tumor immune microenvironment markers (such as PD-L1). Increasing evidence indicates that AI can identify and predict tumor gene mutation status and immune expression on the basis of images. Tan et al. [64] have constructed a stacked ensemble model by fusing one deep learning model and five machine learning models for predicting EGFR and ALK mutation status. The final stacked ensemble yielded an optimal diagnostic performance, thereby providing crucial guidance in TKI selection for targeted therapy in patients with NSCLC. Park et al. [65] have predicted a cytolytic activity score (CytAct) by using a deep learning-based biomarker, which has predicted outcomes in lung adenocarcinoma by estimating a tumor immune profile with FDG-PET noninvasively. Wu et al. [66] have developed an AI system using whole slide images to automatically assess the tumor proportion score of PD-L1 expression. The deep learning model also improves diagnostic repeatability and efficiency for pathologists.

3.3 Treatment response of NSCLC
TNM stage is considered the main prognostic factor after surgery. However, patients with the same stage show significantly different responses to treatment [67]; therefore, the capability to stratify patients according to prognosis is urgently needed, to select those at high risk of recurrence and achieve individualized treatment. In 2020, Wang et al. [68] confirmed that the radiomics label can be used as an independent imaging signature to predict the postoperative recurrence-free survival in patients with early NSCLC and to guide personalized clinical treatment. Kim et al. [69] have developed a CT-based deep learning model to predict disease-free survival, which performs well for patients with clinical stage I lung adenocarcinoma. Adjuvant chemotherapy based on cisplatin is recommended for patients with stage II NSCLC [70]. However, the overall 5-year survival benefit of adjuvant chemotherapy compared with that of surgery alone has been controversial [71, 72]. Because results from different trials have been inconsistent, an accurate biomarker is needed to select patients who may benefit from adjuvant chemotherapy. Vaidya et al. [73] have constructed a nomogram using a radiomics signature that can noninvasively evaluate the survival benefit and predict the actual efficacy of adjuvant chemotherapy.

Preoperative neoadjuvant chemotherapy improves resectability by decreasing tumor size and metastasis. However, the survival benefits of surgery compared with chemoradiotherapy for patients with stage IIIa have long been controversial [74, 75]. Khorrami et al. [76] have retrospectively analyzed 90 patients with stage IIIa disease and selected a total of 13 intratumor and peritumor radiomics features, which can predict pathological remission after neoadjuvant chemotherapy (≤10% residual tumor defined as major pathologic response), with an AUC of 0.90 in the training group and 0.86 in the test group, respectively; the radiomics characteristics
were also associated with overall and progression-free survival. Platinum-based chemotherapy is the standard first-line treatment for advanced NSCLC. However, the objective response rate of the initial treatment regimen was only 24%–31% [77]. Currently, no clinically validated biomarkers can be used to select patients with advanced NSCLC who will benefit from platinum chemotherapy regimens. Khorrami et al. [77] have performed a retrospective study of 125 patients receiving pemetrexed plus platinum chemotherapy. Joint clinical data and intra- and peritumor and radiomics features were analyzed and used to predict lung cancer response to chemotherapy (training group and test group AUC values of 0.82 and 0.77, respectively). The study has indicated that in baseline CT image building, the radiomics signature can be individualized for predicting response to chemotherapy, and is significantly associated with tumor progression and overall survival time.

Before clinical decision-making, evaluation of the therapeutic effects of EGFR-TKIs should be predicted, and priority in alternative treatments should be given to patients at high risk of rapid tumor progression [78]. A multicenter retrospective study has used an imaging radiomics model to predict progression-free survival after EGFR-TKI therapy for patients with stage IV NSCLC with EGFR mutation. A study has included 117 patients treated with EGFR-TKI and selected 12 key characteristics to establish a predictive model. The AUC of 10-month progression-free survival was close to 0.72, and that of 1-year progression-free survival was close to 0.80. The CT-based predictive strategy can predict the DFS probability of EGFR-TKI therapy in NSCLC, thus improving the personalized management of TKIs [79].

AI has a potentially important role in predicting the response to immunotherapy. Vaidya et al. [80] have retrospectively analyzed 109 patients with advanced NSCLC receiving PD-L1/PD-1 inhibitor treatment and divided them into responders and nonresponders. The random forest classifier successfully distinguished excessive progression from other treatment responses. This study suggests that AI may help identify patients at high risk of progression of advanced NSCLC treated with PD-L1/ PD-1 inhibitors. Another study has used PET-CT images to construct a model to select patients who are most likely to benefit from immunotherapy. Therefore, these data can provide accurate and personalized evidence to support clinical treatment decision for patients with advanced NSCLC [81].

A summary of the AI studies in NSCLC can be found in Table 2.

4. LIVER CANCER

Primary liver cancer is the fourth leading cause of cancer-associated deaths worldwide. Because 90% of liver cancer cases are due to cirrhosis, very early liver cancer nodules are difficult to diagnose by visual evaluation, which depends on the diagnostic experience of radiologists and is qualitative. Therefore, developing a highly sensitive, objective and quantitative diagnostic method is urgently needed in clinical practice. AI technology can efficiently use imaging examination results to provide more valuable information for preoperative diagnosis and prediction of treatment response to aid in developing individual treatment plans.

4.1 Diagnosis of liver cancer

In the diagnosis of liver cancer, the accuracy of diagnosis has been significantly improved with AI. Nie [82] has developed CT-based imaging models for the preoperative differentiation of hepatocellular carcinoma (HCC) from hepatocellular adenoma and focal nodular hyperplasia. Each model has good differentiation ability, with an AUC exceeding 0.94. Ponnoprat [83] has developed an image radiomics model based on multiphase CT enhanced images to distinguish HCC from intrahepatic cholangiocarcinoma (ICC), and achieved a model accuracy of 88%. Some researchers have also applied radiomics techniques to the differential diagnosis of HCC subtypes. Liu [84] has reported an MRI imaging radiomics model for distinguishing HCC, ICC and HCC-ICC, with an AUC of 0.77. Huang [85] has constructed a diagnostic model for dual-phenotypic HCC based on GD-Eob-DTPA-enhanced MR images. Currently, CT- or MR-based imaging radiomics techniques have shown good diagnostic performance similar to that of experienced radiologists.

In the diagnosis of liver cancer, many preliminary studies have assessed the potential application value of deep learning in the differential diagnosis of liver cancer. Most of these studies have classified liver lesions into different categories on the basis of MR or CT images and compared them with the diagnostic findings of radiologists with different levels of experience. Yasaka, [86] has developed an algorithm for liver tumor classification based on multiphase CT enhanced images, with an accuracy of 0.84 on the test set. Hamm [87] has used an iteratively optimized CNN network to diagnose six common hepatic focal lesions in multiphase MR enhanced images. The model's accuracy, sensitivity and specificity are 0.92, 0.92 and 0.98, respectively [88]. An interpretable model analysis system has been constructed, and the pretrained deep neural network decision-making principle has been illustrated through analysis of key image features of internal subsets.

4.2 Genetics and molecular marker detection

Many studies on the molecular typing of HCC have shown that radiomics technology can extract complete high-dimensional image information from whole tumor images. The radiomics method supplements traditional genetic analysis, which may have sampling errors, thereby increasing the accuracy of diagnosis. CK19 is a biliary-specific marker associated with poor prognosis of HCC. Wang [89] has used radiomics features...
<table>
<thead>
<tr>
<th>Reference</th>
<th>Application</th>
<th>No. of cases</th>
<th>Imaging modality</th>
<th>Algorithm</th>
<th>Image feature type</th>
<th>Type of validation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maldonado et al. [58]</td>
<td>Prediction of the risk of incidentally detected indeterminate pulmonary nodules</td>
<td>855</td>
<td>LDCT</td>
<td>Brock model, BRODERS classifier</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.87 for the Brock model; AUC of 0.90 for the BRODERS model.</td>
</tr>
<tr>
<td>Hu et al. [60]</td>
<td>Classification of benign versus malignant GGNs</td>
<td>513</td>
<td>CT</td>
<td>3D U-Net and DNN classification model, SVM classifier</td>
<td>Deep learning, radiomics, fusion</td>
<td>Internal validation</td>
<td>Accuracy of 75.6%</td>
</tr>
<tr>
<td>Zhu et al. [61]</td>
<td>Distinguishing SCC from ADC</td>
<td>129</td>
<td>CT</td>
<td>LASSO logistic regression model</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.893</td>
</tr>
<tr>
<td>Park et al. [65]</td>
<td>Prediction of CytAct score</td>
<td>152</td>
<td>FDG-PET</td>
<td>3D CNN model</td>
<td>Deep learning</td>
<td>External validation</td>
<td>Spearman rho = 0.32, p = 0.04 in SNUH cohort; spearman rho = 0.47, p = 0.07 in TCGA cohort</td>
</tr>
<tr>
<td>Wu et al. [66]</td>
<td>Assessment of the TPS of PD-L1 expression</td>
<td>251</td>
<td>WSIs</td>
<td>U-Net structure with residual blocks</td>
<td>Deep learning</td>
<td>External validation</td>
<td>Accuracy of 0.9326; specificity of 0.9641</td>
</tr>
<tr>
<td>Wang et al. [68]</td>
<td>Prediction of RFS in patients with resected stage I NSCLC</td>
<td>378</td>
<td>CT</td>
<td>LASSO logistic regression model</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>Radiomics signature: C-index of 0.776; Radiomics signature integrated with the histology C-index of 0.829</td>
</tr>
<tr>
<td>Kim et al. [69]</td>
<td>Prediction of DFS in patients with lung adenocarcinoma</td>
<td>908</td>
<td>CT</td>
<td>DLPJ model</td>
<td>Deep learning</td>
<td>Internal and external validation</td>
<td>C-index of 0.74–0.80 in the internal validation and 0.71–0.78 in the external validation</td>
</tr>
<tr>
<td>Khorrami et al. [76]</td>
<td>Prediction of MPR</td>
<td>90</td>
<td>CT</td>
<td>Multivariate Cox regression model</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.90±0.025 in the training set and 0.86 in the test set; radiomic signature also significantly associated with OS (HR = 11.18) and DFS (HR = 2.78) in the testing set</td>
</tr>
<tr>
<td>Song et al. [79]</td>
<td>PFS prediction for TKI therapy in multicenter patients with stage IV EGFR-mutated NSCLC</td>
<td>314</td>
<td>CT</td>
<td>LASSO logistic regression model</td>
<td>Radiomics</td>
<td>External validation</td>
<td>C-index of the nomogram of 0.743 for the training cohort and 0.718 and 0.720 for the two validation cohorts, respectively</td>
</tr>
<tr>
<td>Vaidya et al. [80]</td>
<td>Characterization of the response patterns in patients with NSCLC treated with PD-1/PD-L1 inhibitors</td>
<td>109</td>
<td>CT</td>
<td>Random forest classifier</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.85±0.06 in the training set and 0.96 in the validation set</td>
</tr>
</tbody>
</table>

NSCLC, non-small cell lung cancer; GGNs, ground glass nodules; SCC, squamous cell carcinoma; ADC, lung adenocarcinoma; CytAct, cytolytic activity score; WSIs, whole slide images; TPS, tumor proportion score; RFS, recurrence-free survival; DFS, disease-free survival; MPR, major pathological response; PFS, progression-free survival; PD-1, anti-programmed cell death 1; PD-L1, programmed cell death ligand 1.
to predict CK19 in HCC on the basis of gadoxetic acid-enhanced MRI. Combined with radiomics characteristics and clinical factors, the sensitivity and specificity of the validation model are 0.769 and 0.818, respectively.

The Ki-67 index is another molecular biomarker for recurrence and poor prognosis in HCC. Recently, several studies have predicted Ki-67 on the basis of preoperative CT and MRI. Wu [90] has evaluated the correlation between radiomics features and the Ki-67 labeling index. The radiomics parameters predicted Ki-67 status with AUCs ranging from 0.777 to 0.836. Ye [91] has investigated the whole-lesion texture analysis based on preoperative gadoxetic acid-enhanced MRI for predicting Ki-67 status in HCC. The model combining the texture features and clinical data showed high discrimination ability, with an AUC of 0.795.

4.3 Response to treatment
Radiomics methods have been used to evaluate the response to local ablation, recurrence, surgical resection and arterial chemoembolization (TACE) [92].

Wen [93] has built a model to predict the early recurrence of liver cancer after radiofrequency ablation (within 2 years) for local liver ablation. Multivariate logistic regression analysis has established a predictive model including preoperative platelet count and the radiographic Rad-Score, with an AUC of 0.98. Ma [94] has developed an imaging radiomics model based on dynamic contrast-enhanced ultrasound to predict recurrence after ablative therapy in patients with HCC with a diameter of less than 5 cm. The joint imaging radiomics model based on a deep learning reference has shown better prediction for early recurrence, with AUC of 0.89 vs. 0.84. Furthermore, a model combining dynamic contrast-enhanced ultrasound imaging and clinical features has been used to stratify the risk of late relapse.

Zheng [95] has developed programs combining CT imaging radiomics and clinical features to predict recurrence-free survival and overall survival outcomes after hepatectomy for isolated HCC for surgical liver resection. Kong [96] has explored the feasibility of preoperative MRI features to predict the response of liver tumor to TACE. Consequently, eight imaging features associated with TACE response have been screened, and a predictive model has successfully been constructed. Kozzi [24] has performed radiomics analysis on CT-enhanced images of patients with HCC before radiotherapy for liver radiotherapy treatment. Combined with the RECIST criteria, the CT-enhanced image radiomics model for patients with HCC before radiotherapy can be used to evaluate the efficacy of radiotherapy. Chen [97] has applied radiomics analysis of CT-enhanced images of patients with liver cancer before radiotherapy and concluded that the CT-enhanced image radiomics model can evaluate the efficacy of radiotherapy, with an AUC of 0.80.

A summary of the AI studies in liver cancer can be found in Table 3.

5. COLORECTAL CANCER
CRC ranks third among all malignant tumors worldwide [1, 98]. CT and MRI of CRC are crucial for tumor evaluation, which relies heavily on radiologists’ experience, thus affecting diagnostic accuracy and consistency. Recently, AI has provided more precise methods to evaluate CRC.

5.1 Diagnosis of colorectal cancer
For primary tumor (T), lymph node (N) and metastasis (M) staging in CRC, both CT and MRI are essential. AI has proven useful in TNM staging based on CT and MRI. Wang et al. [99] have developed four network models to distinguish among cancer and normal, M0 and M1, normal and elevated carcinoembryonic antigen, and clinical stage I–II and III–IV, on the basis of The Cancer Genome Atlas database. The predictive accuracy in a tenfold cross-validation test was 93.75–99.39%, 80.58–88.24%, 67.21–92.31% and 59.13–68.85%, respectively. You et al. [100] have developed a support vector machine model for T staging diagnosis (T1/T2 vs. T3/T4) using MRI data for 154 patients. In the test set, the models based on T2 WI, ADC and the above two sequences combined achieved AUCs of 0.845, 0.881 and 0.910, respectively. The prediction of lymph node metastasis remains a major topic. Huang et al. [101] have built a radiomics nomogram to predict lymph node status in CRC on the basis of CT of primary tumors, with a C-index of 0.778 in internal validation. Liu et al. [102] have built five support vector machine models to predict LN metastasis of the tumor and mesorectum in rectal cancer. The model based on the combined data, including clinical information, radiomics features from tumor and mesorectal areas achieved the best performance, with an AUC of 0.832. Metastasis is an important risk marker of poor prognosis [103-105]. Li et al. [8] have developed and assessed a radiomics model to predict liver metastasis in CRC on the basis of primary tumor CT data. The model using both the radiomics and clinical information signatures performed best, with an AUC of 0.899 in the test set.

5.2 Histopathological aspects and genetic detection
AI provides a new method to predict the risk factors in CRC that cannot be interpreted by radiologists. Patients with tumor invasion of nerve structures and the lympho-vascular wall, histologically defined as lymphatic vascular infiltration and perineural invasion, respectively, may have poor prognosis. Guo et al. [106] have built a radiomics model based on T2 WI and CE-CT to predict perineural invasion, with an AUC of 0.884 in the testing datasets. Zhang et al. [107] have built a similar radiomics model to predict lymphatic vascular infiltration, with an AUC of 0.876. Microsatellite instability (MSI) is crucial for making individualized treatment plans, and is defined as the loss of one or more mismatch repair proteins.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Application</th>
<th>No. of cases</th>
<th>Imaging modality</th>
<th>Algorithm</th>
<th>Image feature type</th>
<th>Type of validation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis of liver cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al. Nie [82]</td>
<td>Preoperative differentiation of FNH from HCC</td>
<td>156 patients with FNH (n = 55) and HCC (n = 101)</td>
<td>Contrast CT image</td>
<td>Multivariate logistic regression analysis</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.917; 95% CI, 0.800–1.000)</td>
</tr>
<tr>
<td>Ponnoprat et al. [83]</td>
<td>Differentiation between HCC and ICC</td>
<td>187</td>
<td>Contrast CT image</td>
<td>CNN model Support vector machine model</td>
<td>Radiomics</td>
<td>Tenfold cross-validation</td>
<td>Correct detection of 89.30% of HCC and 84.42% of ICC</td>
</tr>
<tr>
<td>Liu et al. [84]</td>
<td>Differentiation between combined HCC-CC vs. CC and HCC</td>
<td>85</td>
<td>MRI and CT</td>
<td>Support vector machine classifier</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.77</td>
</tr>
<tr>
<td>Huang et al. [85]</td>
<td>Description of clinical characteristics and outcomes of DPHCC</td>
<td>50</td>
<td>Gd-EOB-DTPA-enhanced MRI</td>
<td>Multi-layer perceptron, support vector machines, logistic regression and K-nearest neighbor</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>Accuracy of 0.766 in PVP, 0.798 in DP, 0.756 in HBP, 0.798 in ICC</td>
</tr>
<tr>
<td>Yasaka et al. [86]</td>
<td>Differentiation of liver masses</td>
<td>460</td>
<td>Contrast CT image</td>
<td>CNN</td>
<td>Deep learning</td>
<td>Internal validation</td>
<td>Accuracy of 0.84.</td>
</tr>
<tr>
<td><strong>Treatment response</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wen et al. [93]</td>
<td>Preoperative prediction of early recurrence (≤2 years) of small HCC</td>
<td>111</td>
<td>MRI</td>
<td>Multivariable logistic regression</td>
<td>Radiomics</td>
<td>Fivefold cross-validation</td>
<td>AUC of 0.980</td>
</tr>
<tr>
<td>Ma et al. [94]</td>
<td>Prediction of early and late recurrence in patients with a single HCC lesion ≤ 5 cm after thermal ablation</td>
<td>318</td>
<td>Contrast-enhanced ultrasound</td>
<td>Multivariable logistic regression</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>C-index of 0.77</td>
</tr>
<tr>
<td>Zheng et al. [95]</td>
<td>Estimation of postoperative recurrence and survival in patients with solitary HCC</td>
<td>319</td>
<td>Contrast CT image</td>
<td>Cox's hazard regression</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>Recurrence: hazard ratio: 2.472; survival hazard ratio: 1.558</td>
</tr>
<tr>
<td>Kong et al. [96]</td>
<td>Prediction of the response of HCC to TACE</td>
<td>86</td>
<td>MRI</td>
<td>Multivariable logistic regression</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.794</td>
</tr>
<tr>
<td>Chen et al. [97]</td>
<td>Prediction of treatment response to first TACE in patients with intermediate-stage HCC</td>
<td>595</td>
<td>Contrast CT image</td>
<td>Multivariable logistic regression</td>
<td>Radiomics</td>
<td>External validation</td>
<td>AUC of 0.90</td>
</tr>
</tbody>
</table>

FNH, focal nodular hyperplasia; HCC, hepatocellular carcinoma; CC, cholangiocarcinoma; DPHCC, dual-phenotype hepatocellular carcinoma.
Review

5.3 Treatment response of colorectal cancer

Treatment response evaluation in locally advanced or advanced CRC is a critical clinical process [113]. Distinguishing minimal residual tumors after neoadjuvant chemoradiotherapy and predicting treatment effectiveness remain challenges in CRC [114]. Kim et al. [115], in a study at identifying pathological complete response (pCR, defined as ypT0), have shown that a radiomics model based on T2 WI performs better than radiologists (AUC, 0.82 vs. 0.74; sensitivity, 80.0% vs. 15.6%). Several studies [15, 116, 117] have developed radiomic models to predict pCR (defined as ypT0N0), with AUCs of 0.812–0.89 [116, 117]. Furthermore, Tian et al. [15] have constructed a RadioPathomics Integrated Prediction System to predict pCR after neoadjuvant chemoradiotherapy by using MRI and biopsy whole-slide images, which may aid in clinical decision-making.

Adjuvant chemotherapy is an important treatment strategy. Therefore, predicting the response to adjuvant chemotherapy has become a major topic. On the basis of T2 WI and ADC, Tian et al. have built a radiomic model to predict distant metastasis of LARC after adjuvant chemotherapy and shown that a higher radiomic signature indicates poorer prognosis. Patients with pN2 and a low-radiomic signature could receive adjuvant chemotherapy to achieve better prognosis [118]. Hence, radiomics may aid in the formulation of individual treatment plans by screening potential patients who would benefit from therapy.

The AI studies of CRC are summarized in Table 4.

6. CHALLENGES AND FUTURE DIRECTIONS

Imaging plays an essential role in the diagnosis and treatment of tumors. Currently, AI is a major research area. However, many problems are becoming apparent as the number of studies increases.

Insufficient data are a bottleneck in building radiomics and deep learning models, particularly in tumor diagnosis and prognostication. Retrospective clinical data are the most readily available, but substantial biases can arise between model training data and validation data. Constructing multicenter datasets or using public datasets is necessary to obtain homogeneous and unbiased clinical data. In the future, training data may be obtained from multiple institutions through federated learning techniques [119] and sharing models without the need to exchange data.

Poor generalization is an inherent limitation of imaging histology techniques. Radiomics features depend highly on the scanning protocol, ROI selection and feature extraction method. Many factors can lead to high variability in the extracted imaging histology features. However, only stable, reproducible and discriminable models can be generalized in clinical practice. A recent review has proposed several reproducible and generalizable strategies for image histology analysis [120] that can be used to obtain reproducible histological features. In addition, the image histology stability quality score [121], quantitative networks [122] and image biomarker standardization initiatives [123] have gradually become standard for multicenter data applications.

Poor interpretability is a continuing challenge. The issue of interpretability is increasingly important for imaging biomarkers for developing optimal therapy plans. In recent years, the need for interpretable methods has received increasing attention as deep learning techniques continue to evolve. Owing to the complexity of deep learning techniques, millions of parameters are typically optimized during training, data patterns associated with target outputs are automatically extracted, and spurious data correlations may be identified and used, thus decreasing the system’s reliability. Many studies [124, 125] have combined visualization schemes for uncertainty assessment with computer prediction but require imaging experts to validate the visual assessment results. Consequently, interpretable assessments are highly dependent on human judgment and experience, thus making the methodological design and assessment of AI interpretable studies subjective and prone to cognitive bias. In the future, with the gradual integration of AI systems with multidimensional patient information [126], such as clinical indicators, imaging features, molecular pathways and clinical scores, the interpretability of AI models will have more research possibilities.

Automation of the entire AI workflow is another important challenge. Currently, the Picture Archiving and Communication System has been widely used to ensure easy retrieval of data. However, to support further AI research, the data also must be managed for labeling, annotations, segmentations and quality assurance, thus requiring substantial labor. Medical image segmentation based on deep learning has been demonstrated to have good performance in many object segmentation, which can be generally divided into automatic segmentation models and interactive segmentation models [127]. Deploying the constructed segmentation models among radiologists is challenging. The advantage of fully automatic segmentation models is that all segmentation results are provided...
Table 4 | Summary of key studies on the use of artificial intelligence in diagnosis and prediction of treatment response in colorectal cancer.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Application</th>
<th>No. of cases</th>
<th>Imaging modality</th>
<th>Algorithm</th>
<th>Image feature type</th>
<th>Type of validation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al. [99]</td>
<td>Clinical stage I–II/III–IV</td>
<td>633</td>
<td>CT</td>
<td>BP and LVQ</td>
<td>Artificial neural networks</td>
<td>Tenfold cross-validation test</td>
<td>Accuracy of 59.13–68.85%</td>
</tr>
<tr>
<td>Liu et al. [102]</td>
<td>Prediction of lymph node metastasis</td>
<td>186</td>
<td>MRI</td>
<td>SVM</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.832</td>
</tr>
<tr>
<td>Li et al. [8]</td>
<td>Prediction of liver metastasis</td>
<td>100</td>
<td>CT</td>
<td>Logistic regression</td>
<td>Radiomics</td>
<td>Cross-validation set and test set</td>
<td>AUC of 0.899</td>
</tr>
<tr>
<td>Guo et al. [106]</td>
<td>Prediction of preoperative PNI</td>
<td>94</td>
<td>CT and MRI</td>
<td>Multivariate logistic regression analysis</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.884</td>
</tr>
<tr>
<td>Fan et al. [110]</td>
<td>Prediction of MSI</td>
<td>119</td>
<td>CT</td>
<td>Logistic regression model</td>
<td>Radiomics</td>
<td>Internal validation</td>
<td>AUC of 0.752</td>
</tr>
<tr>
<td>He et al. [112]</td>
<td>Prediction of KRAS mutation status</td>
<td>157</td>
<td>CT</td>
<td>ResNet</td>
<td>Deep learning</td>
<td>Internal validation</td>
<td>AUC of 0.90</td>
</tr>
<tr>
<td>Kim et al. [115]</td>
<td>Prediction of pCR</td>
<td>898</td>
<td>MRI</td>
<td>Least absolute shrinkage and selection operator</td>
<td>Radiomics</td>
<td>Tenfold cross-validation</td>
<td>AUC of 0.82</td>
</tr>
<tr>
<td>Tian et al. [15]</td>
<td>Prediction of pCR</td>
<td>1033</td>
<td>MRI</td>
<td>RAdioPathomics Integrated preDiction System</td>
<td>Radiomics</td>
<td>External validation</td>
<td>AUC of 0.81</td>
</tr>
<tr>
<td>Tian et al. [118]</td>
<td>Prediction of adjuvant chemotherapy responders</td>
<td>629</td>
<td>MRI</td>
<td>Least absolute shrinkage and selection operator (LASSO)-Cox regression</td>
<td>Radiomics</td>
<td>External validation</td>
<td>C-indices of 0.803–0.848</td>
</tr>
</tbody>
</table>

PNI, perineural invasion; MSI, microsatellite instability; pCR, pathological complete response.
at one time, thus making this model highly suitable for images with many slices and large volumes of segmentation targets. Although the model uses an advanced attention algorithm to improve the segmentation efficiency [128], the model efficiency may not be satisfactory for the segmentation of multiple and small targets. The interactive segmentation model can automatically identify 2D or 3D regions of interest through the annotation or radiologists clicking on key areas of the image [129]. This model is highly effective for small target organs or ROI annotation with fewer slices but takes more time for images with more slices. In summary, AI segmentation models will help radiologists achieve accurate annotation in most scenarios, and different models should be selected according to task requirements. Regardless of the model used for segmentation with the purpose of accurate quantification, secondary confirmation by a radiologist remains necessary.

AI research continues to deepen, thus bringing hope for practical clinical applications. In addition to glioma, lung cancer, liver cancer and CRC, as described above, AI for practical clinical applications. In addition to glioma, lung cancer, liver cancer and CRC, as described above, AI has numerous applications in breast and cervical cancers [130-132], thus providing substantial advantages for optimization of tumors management strategies.

7. CONCLUSION

Image-based AI will accelerate a revolution in tumor evaluation for diagnosis and prediction of therapy response, and help achieve personalized treatments. The big data era brings hope for AI clinical applications.

ACKNOWLEDGMENT

The authors thank all partners who contributed to this research, including Yu Fu, Zhongping Chen, Fanyang Meng, Kan He and Mingyang Li.

GRANT SUPPORT

China International Medical Exchange Foundation China Medical Imaging International Exchange Special Fund (Z-2014-07-2003-03); Jilin Provincial Department of Finance (Foundation of Jilin Provincial Department of Finance, Establishment of standardized database for CRC and exploration of new diagnosis and treatment model based on large data analysis; JLSZD2019-062); Key Laboratory of Jilin Province, Jilin Provincial Department of Science and Technology (202006010033C); Jilin Provincial Radiation Medical Technology Innovation Center (20190902016TC).

REFERENCES


