An overview of ultrasound-derived radiomics and deep learning in liver

Di Zhang¹, Xian-Ya Zhang², Ya-Yang Duan¹, Christoph F. Dietrich³, Xin-Wu Cui², Chao-Xue Zhang¹

¹Department of Ultrasound, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, ²Department of Medical Ultrasound, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, ³Department of Internal Medicine, Hirslanden Clinic, Bern, Switzerland

Introduction

Liver disease accounts for a significant global burden, with cirrhosis and liver cancer being the eleventh and sixteenth leading causes of mortality, respectively [1]. Medical imaging provides information regarding lesion characteristics in a non-invasive, reproducible manner. The importance of medical imaging in clinical practice has increased significantly from a primary diagnostic tool to a centerpiece contribution to early detection, diagnosis, therapeutic planning, and surveillance of disease [2]. Ultrasound (US) has an essential part in diagnosing liver diseases due to its convenience and non-radiation [3]. However, US examination is highly dependent on the subjective interpretation of images by the operator and lacks quantitative analysis of image features [4]. In addition, traditional US imaging relies purely on human visual interpretations, extracting only the absolute macroscopic features of the disease and losing the more important microbiological information.

Recently, developments in artificial intelligence (AI), especially in radiomics and deep learning, have enabled the extraction of pathophysiology-related information from varied medical imaging and are progressively transforming medical practice. AI applications are extending into domains previously thought to be accessible only to human experts. Recent research has demonstrated that ultrasound-derived radiomics and deep learning represent an enticing opportunity to benefit preoperative evaluation and prognostic monitoring of diffuse and focal liver disease. This review summarizes the application of radiomics and deep learning in ultrasound liver imaging, including identifying focal liver lesions and staging of liver fibrosis, as well as the evaluation of pathobiological properties of malignant tumors and the assessment of recurrence and prognosis. Besides, we identify important hurdles that must be overcome while also discussing the challenges and opportunities of radiomics and deep learning in clinical applications.

Keywords: artificial intelligence; ultrasound; focal liver lesions; radiomics; deep learning

Abstract

Over the past few years, developments in artificial intelligence (AI), especially in radiomics and deep learning, have enabled the extraction of pathophysiology-related information from varied medical imaging and are progressively transforming medical practice. AI applications are extending into domains previously thought to be accessible only to human experts. Recent research has demonstrated that ultrasound-derived radiomics and deep learning represent an enticing opportunity to benefit preoperative evaluation and prognostic monitoring of diffuse and focal liver disease. This review summarizes the application of radiomics and deep learning in ultrasound liver imaging, including identifying focal liver lesions and staging of liver fibrosis, as well as the evaluation of pathobiological properties of malignant tumors and the assessment of recurrence and prognosis. Besides, we identify important hurdles that must be overcome while also discussing the challenges and opportunities of radiomics and deep learning in clinical applications.
Deep learning is a representation learning approach as opposed to traditional machine learning, which enables the original imaging information to be handled directly, automatically extract and select high-dimensional features without using the preset image features [8]. At present, radiomics and deep learning techniques have been implemented to numerous tasks of liver US image analysis, such as traditional diagnostic tasks including the identification of focal liver lesions (FLLs) and staging of liver fibrosis, as well as emerging tasks including the evaluation of pathobiological properties of malignant tumors and prediction of treatment response. In this review, we outline research states of radiomics and deep learning in liver US imaging tasks, while also discussing the challenges and opportunities of radiomics and deep learning in clinical applications.

What is Radiomics?

Overview

Radiomics stems from computational learning theory in pattern recognition, according to which a computational system can develop an algorithm that makes predictions about data by identifying “features” that associate the data with a particular prediction [9,10]. The basic working steps of radiomics include image segmentation, feature extraction and selection, and model building (fig 1).

The application of US-derived radiomics in liver diseases

Diagnosis or grading of liver fibrosis

Liver fibrosis can advance to cirrhosis and hepatocellular carcinoma (HCC). It can ultimately result in death if untreated [11]. Accurate assessment of liver fibrosis is critical for the selection of clinical treatment plans [12]. The standard method for assessing liver fibrosis is still liver biopsy, but it is invasive and can lead to inconclusive results due to sampling errors [13,14]. Al-Hasani et al [15] explored the value of computer extracted B-mode US (BMUS) radiomics features for assessing liver fibrosis using a rat model. They found US images contained a wealth of information that allows for accurate diagnosis of early liver fibrosis using relatively simple machine learning approaches such as logistic regression. The study suggested that the radiomics approach may be a potential tool for noninvasive evaluation of fibrosis. Prospective research by Li et al [16] attempted to further demonstrate the utility of a radiomics models derived from multi-parameter US data for evaluating liver fibrosis in clinical practice. The radiomics signatures derived from original radiofrequency and contrast-enhanced micro-flow imaging provide a better prediction of liver fibrosis staging than traditional radiomics, and the combined model according to the three radiomics signatures has the highest prediction performance with mean area under the receiver operating curve (AUC) value of 0.78-0.85 for distinguishing significant liver fibrosis.

Diagnosis of benign and malignant focal liver lesions (FLLs)

Liver cancer is the 6th most common malignancy and accounts for the third leading cause of cancer-related mortality globally [17]. How to increase the precision of differential diagnosis of benign and malignant liver masses is crucial to reduce misdiagnosis and implement correct therapeutic interventions in a timely manner [18]. Yao et al [19] used radiomics strategy to retrospectively analyze the BMUS, shear wave viscosity imaging, and shear wave elastography (SWE) images of 117 patients with FLLs. AUC of the combined radiomics model on the basis of the three US modalities in differentiating malignant from benign was 0.94. Gatos et al [20] presented a segmentation method for the detection and delineation of FLLs in contrast-enhanced US (CEUS) video sequences, and analyzed the time intensity curve characteristics employing a support vector machine (SVM) classification model, which had an accuracy of 90.3% for the classification of benign and malignant FLLs. Kondo et al [21] proposed another classification method of FLLs derived from multiphase CEUS images. The SVM algorithm is used in combination with the selected optimal features.
to classify the FLLs as benign or malignant with an accuracy of 91.8%.

**Differential diagnosis of liver cancer subtypes**

Preoperative correct recognition of liver cancer subtypes is essential for the rational choice of candidates for liver transplant and hepatectomy, which can improve the overall survival outcome [22,23]. According to the radiomics analysis based on liver US by Peng et al [24], the different histopathological types of primary liver cancer could be identified automatically with an AUC of 0.775 and 0.728 for the discrimination of HCC versus non-HCC and intrahepatic cholangiocarcinoma (ICC) versus combined hepatocellular-cholangiocarcinoma, respectively. Mao et al [25] extracted 1409 radiomic features from US images or derived images of 114 patients, and after feature dimension reduction by LASSO regression, five machine learning classifiers were used to construct a radiomics model. The results indicated that the model based on these five classifiers all had good discriminating ability for primary and metastatic liver cancer, and the performance of logistic regression was superior to that of the other four classifiers, with an AUC reaching 0.816. Qin et al [26] adopted the radiomics strategy to build a prediction model using two-dimensional US to determine the origin of the primary tumor of metastatic liver cancer. After classifying the 254 patients into three subsets according to primary tumor origin, a total of three classification models were developed for each group, which predicted AUCs of 0.767, 0.768, 0.750 for digestive tract versus non-digestive tract tumors, breast cancer versus non-breast cancer, as well as lung cancer versus other malignancies, respectively. The study suggested that radiomics can be used as a complementary approach to determine the source of metastatic liver cancer.

**Predicting histological grade of liver cancer**

The survival rate of patients with liver cancer correlates with the differentiation grade of the tumor [27]. Peng et al [28] used different modelling approaches for radiomics signatures derived from US images of ICC patients and found that the best performing radiomics models for pathological grading was constructed by a hypothetical test combined decision tree. The US-based radiomics model could help physicians to accurately assess the preoperative state of patients, thereby providing favorable decision support for clinical practice.

Multi-modal US-based radiomics algorithms have also been widely employed for the assessment of the histological grade. Wang et al [29] reported that using radiomics strategy to analyze preoperative CEUS images of patients can effectively predict the histological grades of HCC. They selected the most representative image from the BMUS, arterial phase, portal vein phase (PVP) and delay phase (DP) images of each patient, respectively. The radiomics prediction model established after feature extraction and dimensionality reduction, had an AUC of 0.720 for discriminating high grade from low grade, suggesting the potential utility of CEUS-based radiomics for predicting the biological properties of HCC. The research by Dong et al [30] also showed the potential of the radiomic model developed on the basis of the Kupffer phase of Sonazoid CEUS imaging for predicting histological grades in HCC patients.

**Predicting microvascular invasion (MVI) of HCC**

MVI represents a risk variable affecting the operative outcome of individuals with HCC, preoperative identification of its status may assist the surgeon to select an adequate operative approach [31,32]. However, at present, it can only be confirmed by pathological examination after hepatectomy but some researchers have tried to preoperatively identify MVI by US-based radiomics methods [33-35]. In a prospective research, Dong et al [33] applied signal analysis and processing technology to extract radiomics features from original radiofrequency signals data. The radiomics model developed after feature selection and dimension reduction has a prediction accuracy of 92.86% for MVI in HCC patients, which was superior to the radiomics model developed from greyscale US images. Hu et al [34] also constructed a nomogram based on the BMUS radiomics signatures and independent clinical risk variables. The model demonstrated greater discriminatory ability than the predictive model using only clinical risk variables. In another study [35], the relationship between radiomics features based on US image information of peritumoral regions and MVI status of HCC patients was explored. The results indicated that the peritumoral radiomics signature were also potentially valuable for predicting MVI status in patients.

Zhang et al [36] established a CEUS-based radiomic nomogram to assess MVI in HCC patients before surgery. The radiomics nomogram containing a PVP radiomics score and DP radiomics score displayed better discriminative power than the clinical model in both the training and validation sets. The study suggests that CEUS-derived radiomics models can help identify patients at high risk for MVI and have important clinical implications for the development of treatment strategies.

**What is Deep Learning?**

**Overview**

Deep learning refers to machine learning algorithms that perform “feature extraction” and transformation using multiple layers of nonlinear processing [37]. As a subclass of machine learning, deep learning is based on
A deep learning pipeline was developed for ultrasound-derived radiomics and deep learning in liver diseases, which can directly process the original US image, and has the potential to complete various automatic analysis tasks of US images [38]. As the name suggests, it involves using neural networks with many layers stacked. The use of large amounts of layers allows the improved universal approximation properties and the more features to be learned from the data with multiple levels of hierarchy and abstraction [39] (fig 2).

The application of US-derived deep learning in liver diseases

**Diagnosis of fatty liver disease (FLD)**

FLD is a complicated and diverse condition that has a high incidence in the population [40]. In addition to causing decompensated cirrhosis, HCC, and recurrence after liver transplantation, FLD can also affect the progression of other chronic liver diseases [41]. US is the preferred examination approach to evaluate FLD. The ultrasonic image features of the liver, including an increased contrast ratio of liver to kidney and unclear display of intrahepatic duct structure, are suggest the existence of FLD [42,43]. However, the analysis of these characteristics relies on the subjective interpretation by radiologists. A non-invasive and objective diagnostic method is required for the correct diagnosis and early intervention of FLD.

Byra et al [44] established a model based on two-dimensional US images of four liver views to diagnose nonalcoholic FLD and advanced steatosis. The predictive model based on the view of the right posterior portal vein was performed best with an AUC of 0.90 and 0.79 for diagnose nonalcoholic FLD and advanced steatosis, respectively. Han et al [45] developed a one-dimensional convolutional neural network (CNN) based on radiofrequency US data to noninvasively assess nonalcoholic FLD. The accuracy of the CNN algorithm for nonalcoholic FLD in the test group was 96%, and the estimated fat score was correlated with a MRI derived proton density fat fraction with a Pearson r of 0.85 and an average deviation of 0.8%. In a retrospective study [46], only pixels and diagnostic labels of US images were used as input data, and two CNN models trained in advance on ImageNet dataset, including Inception-v3 and VGG-16, were used for feature extraction. The AUC of the model with Inception-v3 and VGG-16 for diagnosing fatty liver were 0.93 and 0.91, respectively. The above studies show that deep learning models are capable of extracting depth information from US images and may be an effective tool for assessing steatosis.

**Diagnosis or grading of liver fibrosis**

CNN models also showed excellent performance for evaluating liver fibrosis. In a multicenter study by Duan et al [47], the authors discussed the diagnostic value of a generative adversarial network based model on BMUS in the classifying fibrosis of S4 and ≥S3. The AUCs of the generative adversarial network model were 0.762 (≥S3), and 0.835 (S4) for external validation cohorts, respectively. Lee et al [48] used 13,608 US images of 3446 patients from one institution for training the CNN prediction model, and 1,232 images of 572 patients were used to validate the classification efficiency of the CNN models for cirrhosis. The accuracy of the CNN model was 76.4% on the external test dataset, and the AUC for assessing cirrhosis was 0.857 on the external test dataset, which was significantly superior to all five radiologists. The CNN model developed by Liu et al [49] can effectively extract the liver capsule signatures from US imaging and was able to evaluate cirrhosis based on the extracted signatures, with the accuracy reaching 0.892.

Wang et al [50] also reported the effectiveness of a deep learning radiomics of elastography (DLRE) strategy in analysis of two-dimensional SWE (2D-SWE) imaging for liver fibrosis non-invasive staging in a multicenter data set. The developed DLRE method had similar diagnostic efficacy to liver biopsy in evaluating cirrhosis and advanced fibrosis, which were substantially better than 2D-SWE as well as other clinical biomarkers. They also found that the performance of DLRE was improved when more 2D-SWE scans were obtained from each participant. The deep learning algorithm can more effectively fuse information from multi-modal US images.
and may provide a viable replacement for invasive liver biopsy.

Transfer learning based on CNN was also developed for liver fibrosis grading. Lu et al [51] constructed a new model named DLRE 2.0 based on the study of Wang et al [50] using transfer learning method. For evaluating of significant fibrosis (≥ F2), AUC of the DLRE2.0 achieved 0.92, which was significantly different from the previous DLRE model (AUC=0.84). The transfer learning model proposed by Xue et al [52] can effectively combined the image data from BMUS and elastogram US and the multi-modal prediction model has higher prediction efficiency compared with the model based on gray-scale or elastogram alone.

**Diagnosis of benign and malignant FLLs**

A deep learning model [53] established with a supervised attention mechanism could simultaneously detected FLLs and distinguish between benign and malignant. Through three times of cross-validation, the average AUC of the model for FLLs detection and distinguishing benign from malignant was 0.935 and 0.916, respectively. Hassan et al [54] proposed a deep learning system with classification accuracy of 97.2% for liver diseases, which is higher than the three machine learning classification methods.

Yang et al [55] conducted a multicenter study to developed a CNN model for classifying FLLs. A total of 24,343 US images from 13 hospitals were incorporated into the data set to develop and validate the CNN model. The combined CNN ModelLBC (lesion + background + clinical) based on lesion, liver background and clinical-sonic factors (such as HBV, HCV, lesion margin, morphology, etc.) had an AUC of 0.924 for the differentiation of benign and malignant FLLs. Moreover, the accuracy of the ModelLBC was similar to the contrast-enhanced CT but slightly lower than contrast-enhanced MRI. This strategy can help inexperienced radiologists improve performance and may have the potential to decrease the dependence of doctors on contrast-enhanced CT/MR and biopsy.

**Assessment of recurrence and prognosis of HCC**

The current treatment methods for liver cancer include surgery, transcatheter arterial chemoembolization (TACE), microwave ablation and so on, achievement of the best therapeutic effectiveness requires careful evaluation of candidates for each therapeutic approach [56-58]. Accurate and personalized prediction of tumor response after treatment has important clinical implications for the overall treatment of patients with liver cancer. Wu et al [59] used the CNN ResNet 18 framework to build a prognostic prediction model for predicting relapse of HCC after microwave ablation based on gray-scale US. The C-index for early relapse, late relapse and relapse-free survival predicted using the model were 0.695, 0.715 and 0.721, respectively. This study shows that analysis of US images using CNN framework is a feasible, and effective method for predicting the prognosis of HCC.

Liu et al [60] designed two deep learning network structures based on CEUS to predict two-year progression-free survival of HCC patients undergoing radiofrequency ablation and surgical resection, respectively. The C-index of the established prediction models for radiofrequency ablation and surgical resection were 0.726 and 0.741, respectively. Moreover, they found that using the predictive model of surgery to predict patients who originally received ablation treatment, 17.3% of ablation patients may have a longer two-year progression-free survival if they undergo surgery. In contrast, the predictive model of ablation predicted that 27.3% of surgical patients may have a longer two-year progression-free survival if they received ablation. Therefore, the CEUS-based models enable accurate assessment of survival of HCC patients and facilitate the selection of optimal treatment. In another study by their group [61], they utilized a deep learning strategy to quantitatively analyses CEUS video. Their group constructed three models to predict the personalized response (objective-response and non-response) of HCC patients after first TACE session, respectively. The AUC of the CEUS model was significantly higher than the other two machine learning models (AUC=0.93, 0.80 and 0.81, respectively). Ma et al [62] analyzed the dynamic CEUS images on the basis of two branch convolution recurrent networks for predicting the recurrence of a single HCC with a diameter ≤ 5cm after thermal ablation. The whole CEUS video was divided into two inputs based on AP and PVP & DP after motion calibration. After the features of each part were combined, the prediction model was established by connecting a 256 hidden unit multilayer perceptron. The prediction model yielded a predicted AUC of 0.84 for early recurrence and a C-index of 0.77 for late recurrence in the test cohort. Deep learning algorithms were shown to be able to effectively utilize the CEUS video information, to enable accurate, personalized predictions.

**Future challenges**

AI is progressively changing the landscape of biomedical research, which offers a significant opportunity for the personalization of healthcare, including a shift from guideline-specific to patient-specific treatment algorithms, providing clinicians and patients with the ability to make data-driven decisions. Radiomics, which is based on imaging methods such as US that noninvasively
provides information on lesion heterogeneity, is currently a highly attractive research tool in the medical field. However, due to the manual segmentation by experts which is required during feature extraction, this time-consuming and challenging process hinders the application of radiomics as a clinical tool in actual clinical diagnosis and treatment. Deep learning, which enables fully automated analysis of images without manual extraction of features, which is the biggest advantage compared with radiomics. However, this advantage is based on the cost of more training data. In contrast to radiomics, the size of the training datasets required for deep learning are typically all on the order of a few thousand or ten thousand cases, ten or more times that of radiomics. However, for some specific problems such as tumor staging or classification, which need to be verified by postoperative pathological data, or for the prediction of survival rate, which need long-term follow-up, the massive collection of such data is very difficult. In addition, deep learning is mostly based on the uninterpretable neural network systems, known as the black box, whose cannot be visualized features may lead to irreversible loss [63]. Interpretability is one of the challenges that deep learning must overcome to be broadly adopted. Furthermore, at the present stage, the development of prediction models based on radiomics and deep learning is still for a specific task, and cannot be used as a substitute for clinicians to comprehensively evaluate multiple abnormalities in the human body.

Notably, a tangible transformation of how promising researches using AI might translate to clinical practice is not yet clear, ethical aspects such as patient privacy protection, data security, diagnosis and treatment security are also problems that radiomics and deep learning need to solve urgently. Will hospitals, governments, and physicians purchase or license the use of these AI-based algorithms? Who is responsible for those patients that are misdiagnosed by AI-based diagnostics? Furthermore, the involvement of AI will definitely modify the traditional doctor-patient relationship, and sonographers will have to adapt to the new role and maintain a positive interaction with patients.

In addition, multi-omics studies encompassing genomics, transcriptomics, proteomics, and metabolomics, have emerged as a hot topic in characterizing the molecular biology of tumors. By correlating multi-omics data with radiomics features obtained from medical images in multiple dimensions, it is possible to truly combine macroscopic and microscopic information to improve disease diagnosis, noninvasive prediction at the molecular level. In the future, combining US radiomics with clinical data and microscopic data to conduct multi-omics studies thus facilitating accurate diagnosis and prediction of biological behavior of liver diseases is a direction worth exploring.

Conclusions

AI applications are extending into domains previously thought to be accessible only to human experts. Thanks to recent advancements in digital data capture and computing infrastructures, the US-derived radiomics and deep learning represent an enticing opportunity to benefit preoperative evaluation and prognostic monitoring of diffuse and focal liver disease. However, as two emerging research approaches, there are still many algorithmic and data standardization issues that need to be further explored. Cross-disciplinary cooperation is required to move the subject forward, and the application of all-encompassing, highly computational clinical models is expected in the coming future. In addition, the ethical issues and challenges arising from the application and development of AI in clinical practice must be taken into account.

Conflict of interest: none

Reference

40. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. JAMA 2015;313:2263-2273.