Role of radiomics in staging liver fibrosis: a meta-analysis

Xiao-min Wang[1]* and Xiao-jing Zhang[2]

Abstract

Background Fibrosis has important pathoetiological and prognostic roles in chronic liver disease. This study evaluates the role of radiomics in staging liver fibrosis.

Method After literature search in electronic databases (Embase, Ovid, Science Direct, Springer, and Web of Science), studies were selected by following precise eligibility criteria. The quality of included studies was assessed, and meta-analyses were performed to achieve pooled estimates of area under receiver-operator curve (AUROC), accuracy, sensitivity, and specificity of radiomics in staging liver fibrosis compared to histopathology.

Results Fifteen studies (3718 patients; age 47 years [95% confidence interval (CI): 42, 53]; 69% [95% CI: 65, 73] males) were included. AUROC values of radiomics for detecting significant fibrosis (F2-4), advanced fibrosis (F3-4), and cirrhosis (F4) were 0.91 [95%CI: 0.89, 0.94], 0.92 [95%CI: 0.90, 0.95], and 0.94 [95%CI: 0.93, 0.96] in training cohorts and 0.89 [95%CI: 0.83, 0.91], 0.89 [95%CI: 0.83, 0.94], and 0.93 [95%CI: 0.91, 0.95] in validation cohorts, respectively. For diagnosing significant fibrosis, advanced fibrosis, and cirrhosis the sensitivity of radiomics was 84.0% [95%CI: 76.1, 91.9], 86.9% [95%CI: 76.8, 97.0], and 92.7% [95%CI: 89.7, 95.7] in training cohorts, and 75.6% [95%CI: 67.7, 83.5], 80.0% [95%CI: 70.7, 89.3], and 92.0% [95%CI: 87.8, 96.1] in validation cohorts, respectively. Respective specificity was 88.6% [95% CI: 83.0, 94.2], 88.4% [95% CI: 81.9, 94.8], and 91.1% [95% CI: 86.8, 95.5] in training cohorts, and 86.8% [95% CI: 83.3, 90.3], 94.0% [95% CI: 89.5, 98.4], and 88.3% [95% CI: 84.4, 92.2] in validation cohorts. Limitations included use of several methods for feature selection and classification, less availability of studies evaluating a particular radiological modality, lack of a direct comparison between radiology and radiomics, and lack of external validation.

Conclusion Although radiomics offers good diagnostic accuracy in detecting liver fibrosis, its role in clinical practice is not as clear at present due to comparability and validation constraints.

Keywords Radiomics, Liver, Fibrosis, Cirrhosis, MRI, CT, Ultrasound

Introduction

Chronic liver disease is an important health concern due to high prevalence of metabolic dysfunction associated fatty liver disease (MAFLD), hepatitis B/C, and alcoholic liver disease. Whereas mortality due to liver disease has declined in some countries like the USA and China, some countries such as India and Mongolia still have higher mortality rates. Increasing trends in the prevalence are also noted in the United Kingdom and Russia [1]. In the USA, 1.8% of the population has a liver disease diagnosis [2]. In China, although the mortality rates have decreased, the prevalence of liver disease is increasing [3]. Globally, the incidence of non-alcoholic steatohepatitis-caused cirrhosis is increasing by 1.35% each year [4].

Liver fibrosis is a modifiable factor that is associated with worse health outcomes, transplants, and mortality

---

*Correspondence:
Xiao-min Wang
wangxiaomin_tmu@126.com
[1] School of Medical Imaging, Tianjin Medical University, No.1, Guangdong Road, Hexi District, Tianjin 300203, China
[2] Department of Radiology, The First Medical Center, Chinese PLA General Hospital, Beijing 100853, China

[5]. Liver fibrosis may develop due to chronic viral infection, long-term alcohol use, or steatohepatitis. It is estimated that liver fibrosis affects 7.7% of individuals in the general population of the United States of America and obese individuals are at much higher risk [6]. The latest stage of fibrosis, cirrhosis, is the eleventh leading cause of mortality [1]. An increasing trend in mortality due to cirrhosis has been observed globally from 1990 to 2017 [7]. Several methods are available to measure liver fibrosis. The liver biopsy is considered diagnostic "gold standard" for staging liver fibrosis. The biopsy is a highly valuable tool but can be associated with complications, sampling errors, and between-observer variations in judgments. Fibrosis can be patchy so that not all parts of the liver contain fibrosis evenly, therefore, a biopsy may fail to capture some samples [8]. Among other methods, serum biomarkers may also help in detecting fibrosis. Radiological methods including magnetic resonance imaging (MRI) magnetic resonance elastography (MRE), computed tomography (CT), ultrasonography, and elastography also provide non-invasive means of measuring fibrosis [9].

Radiomics is a post-radiology process of high-throughput extraction of features from radiological images for conversion into mineable data involving complex processes of artificial intelligence such as machine learning, deep learning, and convolutional neural networks to maximize predictability. It is developed on the premise that pathophysiological tissues and organs contain much information that can be quantified and differentiated from normal tissues and organs. Extraction of a large number of features to form a database and then mining the data for analyses aids decision support leading to improved diagnostic accuracy and prognostic capability [10–12]. Radiomics takes texture as a spatial arrangement of predefined voxels through which complicated features of the image can be read and mathematical calculations of these arrangement characteristics differentiate normal from abnormal. The heterogeneity in the selected features reflects the heterogeneity of histopathological changes [13]. Radiomics features can be morphological, histogram, textural, and high-order features. Morphological features include the shape, size, and volume of the region of interest. Histogram is the plotting of pixel values against pixel frequency and can be used to describe many features such as magnitude, dispersion, asymmetry, peakedness, flatness, randomness, uniformity, etc. Textural features provide spatial relationships between neighboring pixels. High-order features are those acquired after applying filters to images [12, 14].

Several reviews have described the role of radiomics in the diagnosis and staging of various types of cancers [15, 16]. Among other clinical applications, radiomics has been found to be a valuable aid in cardiomyopathy [17, 18], musculoskeletal diseases [19, 20], neurological and psychiatric disorders [21–24], and liver diseases [25]. Several studies have reported the diagnostic accuracy indices of radiomics in diagnosing and staging liver fibrosis. However, there is no synthesis of these outcomes which are sometimes variable and even inconsistent. The present study aimed to evaluate the role of radiomics in diagnosing and staging liver fibrosis by conducting a systematic review of relevant studies and performing meta-analyses of statistical indices.

Method
The present study was conducted by following PRISMA guidelines.

Inclusion and exclusion criteria
A study was included in the meta-analysis if a) it prospectively or retrospectively recruited patients with chronic liver disease who had histologically confirmed fibrosis in the liver; b) performed radiomic analyses based on any radiological modality to diagnose and/or differentiate fibrosis; and c) reported diagnostic accuracy indices of radiomics in diagnosing and differentiating liver fibrosis stages by in comparison with histopathology. Studies were excluded based on the following criteria: a study a) reported diagnostic performance of radiomics for liver fibrosis without adequate statistical data; b) reported the outcomes of pediatric patients; and c) reported the diagnostic accuracy of a combined clinical-radiological radiomics model.

Literature search
Electronic scientific databases (Embase, Ovid, Science Direct, Springer, and Web of Science) were searched for the identification of relevant studies using area-specific keywords. The primary search strategy was "Radiomics AND liver fibrosis OR cirrhosis AND diagnostic accuracy". Secondary keywords were used in several other combinations with this primary string. The detailed literature search strategy is given in Appendix S1. After the identification of studies, reference lists of related articles were also screened for additional studies. The literature search encompassed peer-reviewed research articles published in English from the date of database inception till May 2023.

Data analysis
Data on the design and conduct of studies, patient demographics, clinical characteristics, fibrosis stage, radiomics design and analyses, and diagnostic accuracy outcome data were extracted from the research articles of respective studies and organized in data sheets. The quality

of the included studies was assessed with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) scale. This scale assesses the quality of studies under the domains of risk of bias and applicability concerns by evaluating patient selection, index test, reference standard, and flow and timing. Diagnostic accuracy endpoints (Accuracy, area under receiver operator curve (AUROC), sensitivity, and specificity of various radiomics models compared to histologically proven fibrosis were extracted from the research articles of respective studies and pooled under random effects model using the point estimates and their 95% confidence intervals of these indices. Subgroup analyses were performed with respect to fibrosis stage (significant fibrosis; stages F2-4, advanced fibrosis; stages F3-4, and cirrhosis; stage F4) and with respect to the study cohort (training, test/

validation). Statistical analyses were performed with Stata software (Stata Corporation, College Station, Texas, USA).

Results

Fifteen studies [13, 26–39] were included that were published between 2018 and 2023 (Fig. 1). In these studies, radiomic analyses were performed involving 3718 patients with chronic liver diseases including hepatitis and MAFLD. The age of these patients was 47.3 years [95% confidence interval (CI): 42.0, 52.5]. The proportion of males was 69% [95% CI: 65, 73]. Histologically (biopsy/surgical) confirmed fibrosis stage was F0 in 13% [95% CI: 7,20], F1 in 17% [95% CI: 13, 23], F2 in 21% [95% CI: 17,26], F3 in 16% [95% CI: 13,19], and F4 in 27% [95% CI: 23, 33] of the patients. Radiological
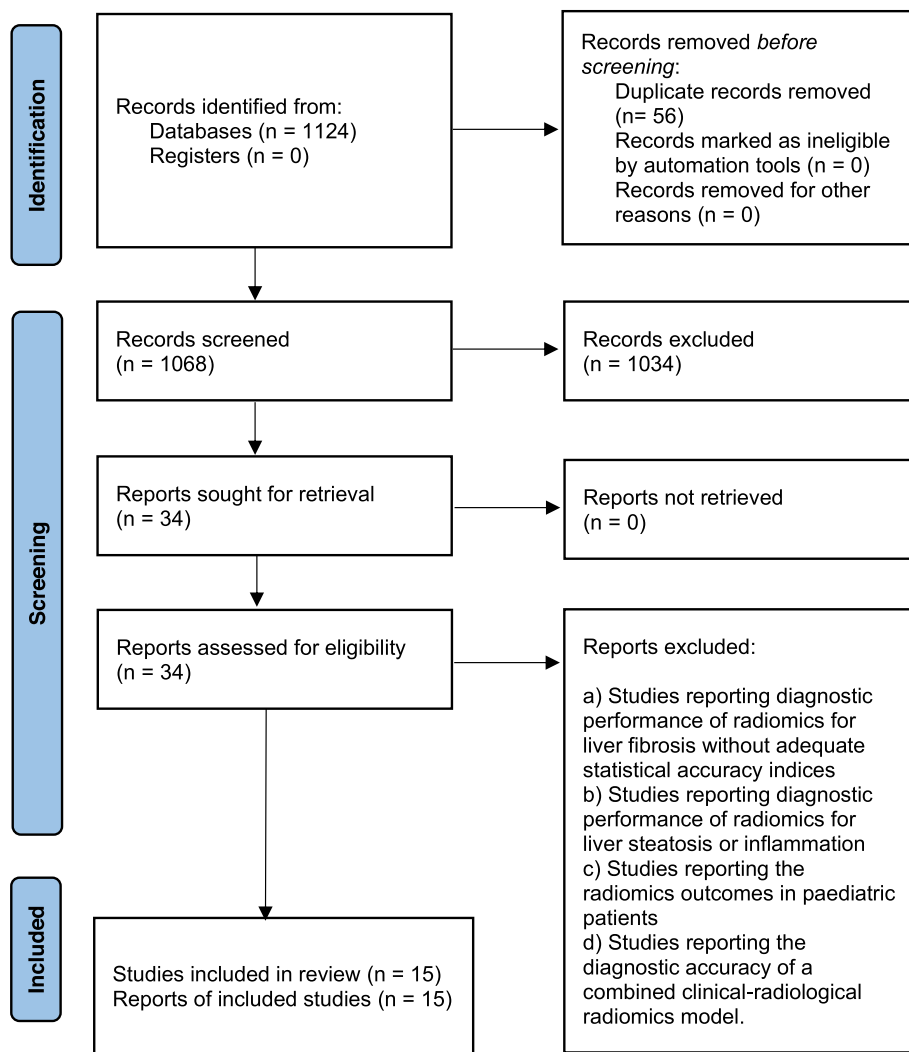

Fig. 1 A flowchart of study screening and selection process

modalities used in these studies were: MRI (5), CT (3), ultrasonography (2), positron emission tomography (1), shear-wave elastography (2), and MRE (2).

Important characteristics of the included studies are given in Table S1. The quality of the included studies was moderate in general according to the QUADAS-2 scale (Table S2). The risk of bias was limited to the retrospective design of studies which could have introduced patient selection bias. Moreover, the interval between the index test and reference test was up to six months which might have impacted adequate flow and timing. Included studies also varied with regards to validation ranging from no validation [26], Five/Ten-fold cross-validation [27, 30, 32, 36, 39], and leave-one-out cross-validation [13] to internal validation [28, 30, 31, 33, 34, 37, 38], and internal and external validation [29]. There was no significant publication bias according to Egger's test (Bias coefficient 0.114 [-11.80, 12.03]; *p*=0.984) or Begg's test (Adjusted Kendall's score -29±18; *p*=0.125) (Figure S1a and b).

The pooled AUROC value of radiomics for the diagnosis of any liver fibrosis was 0.878 [95% CI: 0.850, 0.906]. The AUROC values for the detection of significant fibrosis, advanced fibrosis, and cirrhosis were 0.914 [95% CI: 0.889, 0.938], 0.924 [95% CI: 0.901, 0.946], and 0.944 [95% CI: 0.929, 0.963] respectively in training cohorts, and 0.886 [95% CI: 0.826, 0.905], 0.887 [95% CI: 0.834, 0.939], and 0.930 [95% CI: 0.905, 0.954] respectively in test/validation cohorts (Figs. 2 and 3).

The pooled accuracy of radiomics in diagnosing any liver fibrosis was 83.5% [95% CI: 81.7, 85.4]. The accuracy values of radiomics in the detection of significant fibrosis, advanced fibrosis, and cirrhosis were 80.6% [95% CI: 76.2, 85.1], 83.5% [95% CI: 81.3, 85.8], and 81.6% [95% CI: 78.7, 84.5] respectively in training cohorts, and 77.0% [95% CI: 68.6, 85.5], 84.3% [95% CI: 79.8, 88.8], and 81.3% [95% CI: 77.0, 85.7] respectively in test/validation cohorts (Figures S2a and b).

In training cohorts, the sensitivity of radiomics in diagnosing significant fibrosis, advanced fibrosis, and cirrhosis was 84.0% [95% CI: 76.1, 91.9], 86.9% [95% CI: 76.8, 97.0], and 92.7% [95% CI: 89.7, 95.7] respectively (Fig. 4). In test/validation cohorts, the sensitivity of radiomics in diagnosing significant fibrosis, advanced fibrosis, and cirrhosis was 75.6% [95% CI: 67.7, 83.5], 80.0% [95% CI: 70.8, 89.3], and 92.0% [95% CI: 87.8, 96.1] respectively (Fig. 5).

In training cohorts, the specificity of radiomics in diagnosing significant fibrosis, advanced fibrosis, and cirrhosis was 88.6% [95% CI: 83.0, 94.2], 88.4% [95% CI: 81.9, 94.8], and 91.1% [95% CI: 86.8, 95.5], whereas in test/validation cohorts, the specificity was 86.8% [95%

CI: 83.3, 90.3], 94.0% [95% CI: 89.5, 98.4], and 88.3% [95% CI: 84.4, 92.2] respectively (Figure S3a and b).

Among the included studies, Hu et al. [27] who found thin-sliced CT images to yield better performance of radiomics than thick-sliced CT images reported AUROC values of 0.90 [95% CI: 0.84, 0.96] for F1 vs F2-4, 0.85 [95% CI: 0.78, 0.92] for F1-2 vs F3-4, and 0.94 [95% CI: 0.89, 0.97] for F1-3 vs F4 stages. Lan et al. [13] who studied MRE radiomics found AUROC values of 0.89 [95% CI: 0.84, 0.94] for F0 vs F1-4, 0.93 [95% CI: 0.89, 0.98] for F0-1 vs F2-4, 0.92 [95% CI: 0.88, 0.97] for F0-2 vs F3-4, and 0.95 [95% CI: 0.90, 0.997] for F0-3 vs F4 stages in their echo-planar images. They observed AUROC values of 0.89 [95% CI: 0.83, 0.94], 0.87 [95% CI: 0.81, 0.93], 0.89 [95% CI: 0.84, 0.95], and 0.94 [95% CI: 0.89, 0.997] for F0 vs F1-4, F0-1 vs F2-4, F0-2 vs F3-4 and F0-3 vs F4 stages respectively in their gradient recalled echo images.

Discussion

This meta-analysis found that radiomics exhibits high accuracy in diagnosing and staging liver fibrosis. The AUROC values for the detection of significant fibrosis, advanced fibrosis, and cirrhosis were approximately 0.91, 0.92, and 0.94 in training cohorts and 0.89, 0.89, and 0.93 in validation cohorts, respectively. However, despite the good efficiency of radiomics in diagnosis and staging liver fibrosis observed herein, several factors make these findings inconclusive and dependent on future studies to refine this evidence. For example, several types of radiological modalities were used for radiomics in individual studies and a direct comparison of diagnostic performance between radiomics and radiology was mostly lacking. Moreover, various models were utilized for radiomics, and these studies lacked external validation.

Some studies that could not be included in this meta-analysis have also reported variable diagnostic performance of radiomics in staging liver fibrosis. Cui et al. [40] utilized multiphase CT-based radiomics to stage liver fibrosis and found the sensitivity of diagnosing significant fibrosis, advanced fibrosis, and cirrhosis to be 30-41%, 59-71%, and 84-87%, and the specificities being 84-90%, 71-79%, and 50-58% respectively. Duan et al. [41] observed better performance of ultrasound-based radiomics in diagnosing cirrhosis (AUROC 0.86) than advanced fibrosis (AUROC 0.77). Zhou et al. [42], also reported higher efficieny of ultrasound-based radiomics in diagnosing cirrhosis (AUROC 0.83-0.86) than significant fibrosis (AUROC 0.69-0.71) or advanced fibrosis (AUROC 0.67-0.72).

Many studies have shown that the diagnostic accuracy of radiomics is better than serological biomarkers [30, 42–44]. Sim et al. [32] found better diagnostic performance of MRE radiomics (AUROC 0.97 [95% CI: 0.93, 1])

Study	Selection	AUROC (95% CI)	% Weight
Significant fibrosis			
Lu 2021	DLRE	0.84 (0.82, 0.86)	8.15
Lu 2021	DLRE2.0_C	0.93 (0.91, 0.95)	8.15
Lu 2021	DLRE2.0_CSX	0.96 (0.94, 0.98)	8.15
Park 2019	RFI/LR	0.91 (0.87, 0.94)	7.33
Wang 2019	DLRE	0.99 (0.97, 1.00)	8.35
Wang 2022	LR	0.90 (0.87, 0.94)	7.11
Xue 2020	TL	0.99 (0.97, 1.00)	8.38
Xue 2020	GM non-TL	0.99 (0.98, 0.99)	8.55
Xue 2020	EM non-TL	0.99 (0.98, 0.99)	8.59
Yin 2022	LASSO/LR	0.95 (0.93, 0.98)	7.91
Yin 2022	Adaboost	0.74 (0.67, 0.80)	5.37
Yin 2022	Gradiant boosting	0.85 (0.81, 0.92)	6.02
Yin 2022	XGBoosting	0.78 (0.70, 0.87)	4.24
Zhang 2023	LASSO/LR	0.71 (0.62, 0.81)	3.67
Subtotal (I-squared = 96.6%, p = 0.000)		0.91 (0.89, 0.94)	100.00
.			
Advanced fibrosis			
Park 2019	RFI/LR	0.88 (0.85, 0.92)	9.71
Wang 2022	LR	0.91 (0.88, 0.94)	10.16
Wang 2019	DLRE	0.99 (0.97, 1.00)	11.98
Xue 2020	TL	0.99 (0.98, 1.00)	12.45
Xue 2020	GM non-TL	0.98 (0.97, 0.99)	12.24
Xue 2020	EM non-TL	0.99 (0.98, 0.99)	12.51
Yin 2022	LASSO/LR	0.92 (0.88, 0.95)	9.71
Yin 2022	Adaboost	0.72 (0.63, 0.79)	4.91
Yin 2022	Gradiant boosting	0.83 (0.78, 0.89)	7.24
Yin 2022	XGBoosting	0.79 (0.71, 0.89)	4.22
Zhang 2023	LASSO/LR	0.88 (0.80, 0.96)	4.87
Subtotal (I-squared = 94.0%, p = 0.000)		0.92 (0.90, 0.95)	100.00
.			
Cirrhosis			
Lu 2021	DLRE	0.95 (0.93, 0.97)	8.82
Lu 2021	DLRE2.0_C	0.97 (0.95, 0.99)	8.82
Lu 2021	DLRE2.0_CSX	0.99 (0.97, 1.00)	9.28
Park 2019	RFI/LR	0.87 (0.82, 0.91)	6.05
Wang 2019	DLRE	0.99 (0.98, 1.00)	9.64
Wang 2022	LR	0.84 (0.80, 0.89)	6.10
Xue 2020	TL	0.99 (0.94, 1.00)	8.13
Xue 2020	GM non-TL	0.96 (0.94, 0.97)	9.54
Xue 2020	EM non-TL	0.99 (0.98, 1.00)	9.83
Yin 2022	LASSO/LR	0.91 (0.86, 0.95)	6.05
Yin 2022	Adaboost	0.82 (0.73, 0.88)	3.57
Yin 2022	Gradiant boosting	0.88 (0.78, 0.93)	3.57
Yin 2022	XGBoosting	0.90 (0.83, 0.95)	4.64
Zhang 2023	LASSO/LR	0.94 (0.89, 0.99)	5.95
Subtotal (I-squared = 91.1%, p = 0.000)		0.94 (0.93, 0.96)	100.00

NOTE: Weights are from random effects analysis

Fig. 2 Forest graphs showing the outcomes of meta-analysis of AUROC values of radiomics in diagnosing fibrosis stages in training cohorts

than clinical features (AUROC 0.91 [95% CI: 0.81, 0.97]) in diagnosing significant fibrosis. Wang et al. [34] also reported that the AUROC values for significant fibrosis, advanced fibrosis, and cirrhosis were higher with CT-based radiomics (AUROC 0.88, 0.90, and 0.86) than with either aspartate transaminase-to-platelet ratio (AUROC 0.69, 0.67, and 0.65) or Fibrosis-4 index (AUROC 0.71, 0.71, and 0.7) respectively. Xue et al. [35] also found

better diagnostic performance of multimodal ultrasound radiomics (AUROC 0.9-0.93) than either aspartate transaminase-to-platelet ratio (AUROC 0.72-0.78) or Fibrosis-4 index (AUROC 0.69-0.75) in staging liver fibrosis. Zhao et al. [38] found better diagnostic efficiency of MRI radiomics (accuracy 0.8) than clinical markers (accuracy 0.68) in differentiating non-significant fibrosis from clinically significant fibrosis in the test cohort.

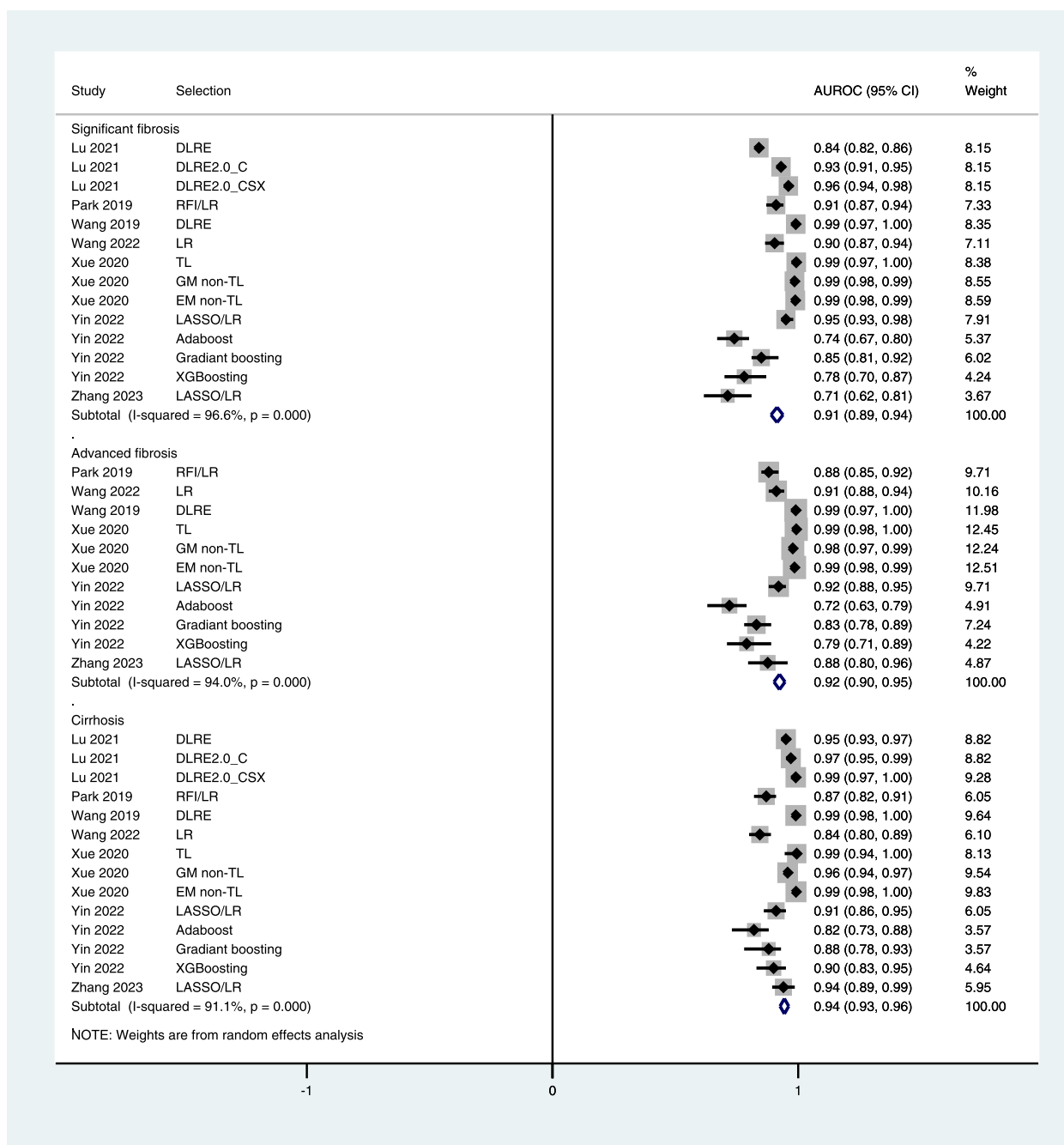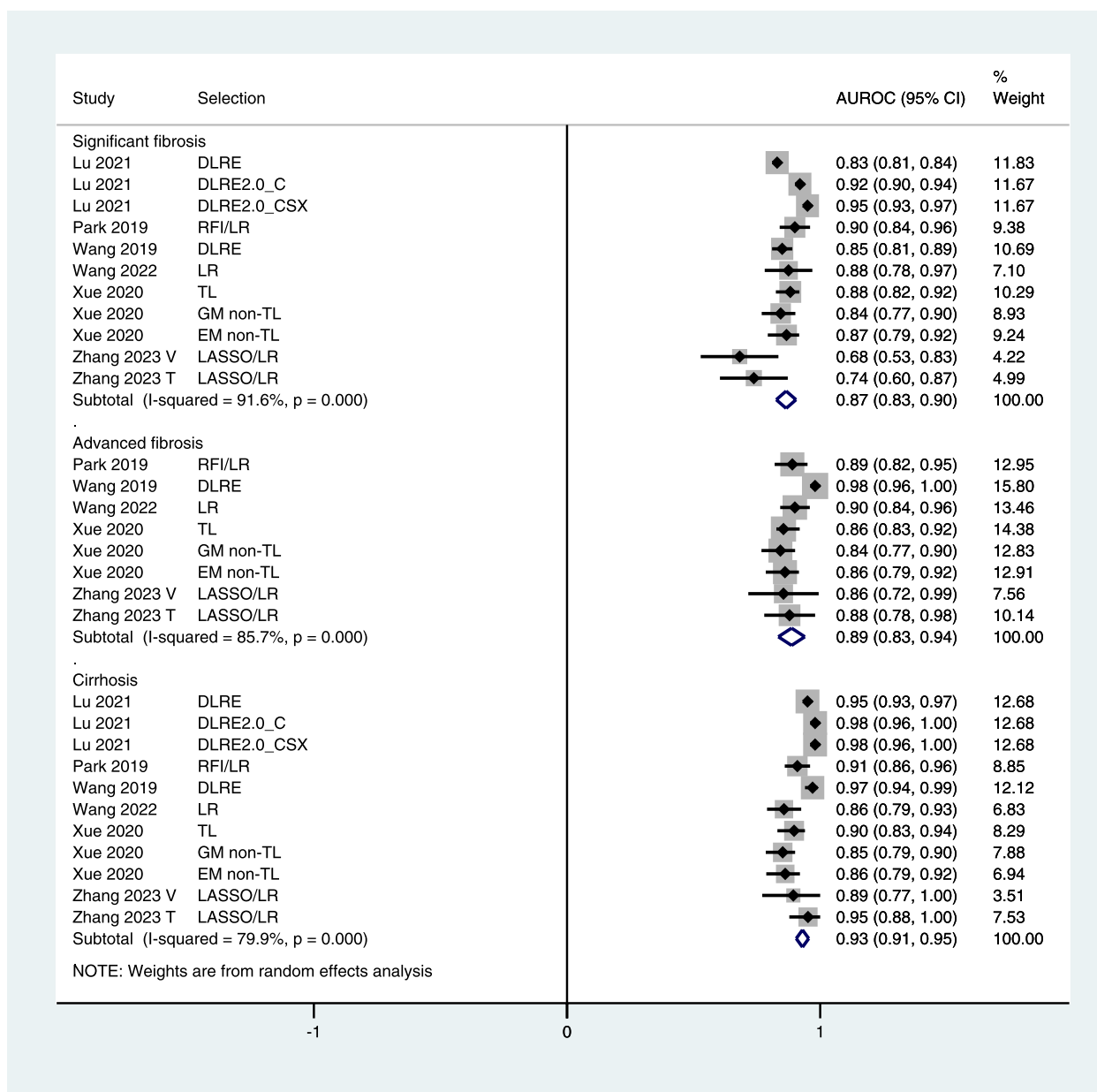| Study	Selection	AUROC (95% CI)	% Weight
Significant fibrosis			
Lu 2021	DLRE	0.83 (0.81, 0.84)	11.83
Lu 2021	DLRE2.0_C	0.92 (0.90, 0.94)	11.67
Lu 2021	DLRE2.0_CSX	0.95 (0.93, 0.97)	11.67
Park 2019	RFI/LR	0.90 (0.84, 0.96)	9.38
Wang 2019	DLRE	0.85 (0.81, 0.89)	10.69
Wang 2022	LR	0.88 (0.78, 0.97)	7.10
Xue 2020	TL	0.88 (0.82, 0.92)	10.29
Xue 2020	GM non-TL	0.84 (0.77, 0.90)	8.93
Xue 2020	EM non-TL	0.87 (0.79, 0.92)	9.24
Zhang 2023 V	LASSO/LR	0.68 (0.53, 0.83)	4.22
Zhang 2023 T	LASSO/LR	0.74 (0.60, 0.87)	4.99
Subtotal (I-squared = 91.6%, p = 0.000)		0.87 (0.83, 0.90)	100.00
.			
Advanced fibrosis			
Park 2019	RFI/LR	0.89 (0.82, 0.95)	12.95
Wang 2019	DLRE	0.98 (0.96, 1.00)	15.80
Wang 2022	LR	0.90 (0.84, 0.96)	13.46
Xue 2020	TL	0.86 (0.83, 0.92)	14.38
Xue 2020	GM non-TL	0.84 (0.77, 0.90)	12.83
Xue 2020	EM non-TL	0.86 (0.79, 0.92)	12.91
Zhang 2023 V	LASSO/LR	0.86 (0.72, 0.99)	7.56
Zhang 2023 T	LASSO/LR	0.88 (0.78, 0.98)	10.14
Subtotal (I-squared = 85.7%, p = 0.000)		0.89 (0.83, 0.94)	100.00
.			
Cirrhosis			
Lu 2021	DLRE	0.95 (0.93, 0.97)	12.68
Lu 2021	DLRE2.0_C	0.98 (0.96, 1.00)	12.68
Lu 2021	DLRE2.0_CSX	0.98 (0.96, 1.00)	12.68
Park 2019	RFI/LR	0.91 (0.86, 0.96)	8.85
Wang 2019	DLRE	0.97 (0.94, 0.99)	12.12
Wang 2022	LR	0.86 (0.79, 0.93)	6.83
Xue 2020	TL	0.90 (0.83, 0.94)	8.29
Xue 2020	GM non-TL	0.85 (0.79, 0.90)	7.88
Xue 2020	EM non-TL	0.86 (0.79, 0.92)	6.94
Zhang 2023 V	LASSO/LR	0.89 (0.77, 1.00)	3.51
Zhang 2023 T	LASSO/LR	0.95 (0.88, 1.00)	7.53
Subtotal (I-squared = 79.9%, p = 0.000)		0.93 (0.91, 0.95)	100.00

NOTE: Weights are from random effects analysis

Fig. 3 Forest graphs showing the outcomes of meta-analysis of AUROC values of radiomics in diagnosing fibrosis stages in test/validation cohorts

Some authors have suggested that a model combining radiomics and clinical biomarkers may further improve the diagnostic accuracy of fibrosis [38, 45].

Despite good diagnostic accuracy values observed for radiomics in liver fibrosis assessment in the present study, inconsistencies were observed in the outcomes of individual studies. Whereas Wang et al. [34] and Lan et al. [13] found radiomics better than radiological elastography in diagnosing advanced fibrosis and cirrhosis, Sim et al. [32] did not find a statistically significant difference. Lu et al. [29] and Sim et al. [32] found that radiomics distinguished well between significant fibrosis (F2-4) and non-significant fibrosis (F0-1). However, Zhang et al. [37] reported that radiomics was unable to distinguish between non-significant fibrosis and significant fibrosis. Zhao et al. [38] reported that a combined use of radiomics and clinical biomarkers performed better than radiomics alone, whereas Sim et al. [32] did not find a significant difference in performance between radiomics and combined use of radiomics, radiology, and clinical biomarkers.
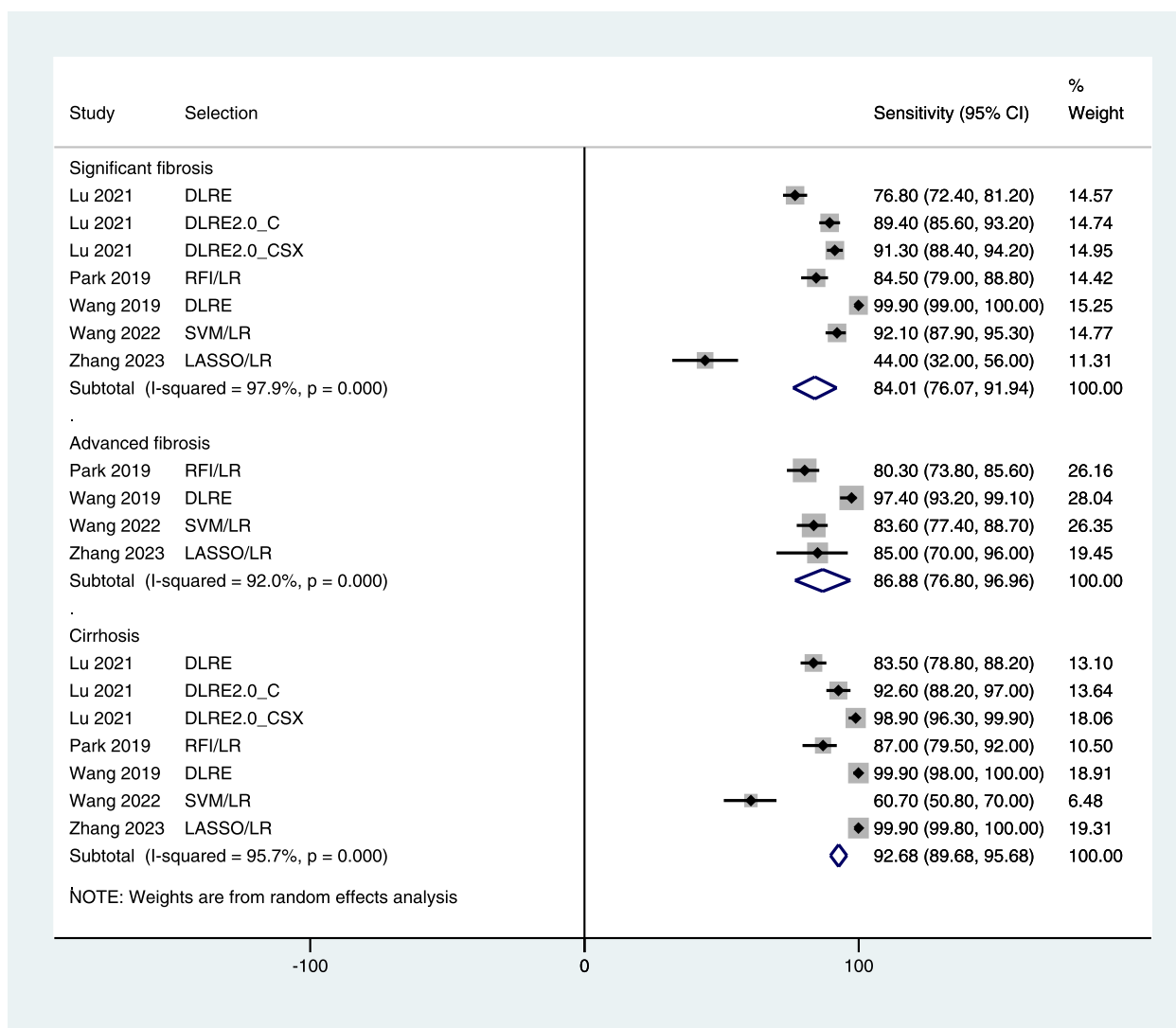
Study	Selection	Sensitivity (95% CI)	% Weight
Significant fibrosis			
Lu 2021	DLRE	76.80 (72.40, 81.20)	14.57
Lu 2021	DLRE2.0_C	89.40 (85.60, 93.20)	14.74
Lu 2021	DLRE2.0_CSX	91.30 (88.40, 94.20)	14.95
Park 2019	RFI/LR	84.50 (79.00, 88.80)	14.42
Wang 2019	DLRE	99.90 (99.00, 100.00)	15.25
Wang 2022	SVM/LR	92.10 (87.90, 95.30)	14.77
Zhang 2023	LASSO/LR	44.00 (32.00, 56.00)	11.31
Subtotal (I-squared = 97.9%, p = 0.000)		84.01 (76.07, 91.94)	100.00
.			
Advanced fibrosis			
Park 2019	RFI/LR	80.30 (73.80, 85.60)	26.16
Wang 2019	DLRE	97.40 (93.20, 99.10)	28.04
Wang 2022	SVM/LR	83.60 (77.40, 88.70)	26.35
Zhang 2023	LASSO/LR	85.00 (70.00, 96.00)	19.45
Subtotal (I-squared = 92.0%, p = 0.000)		86.88 (76.80, 96.96)	100.00
.			
Cirrhosis			
Lu 2021	DLRE	83.50 (78.80, 88.20)	13.10
Lu 2021	DLRE2.0_C	92.60 (88.20, 97.00)	13.64
Lu 2021	DLRE2.0_CSX	98.90 (96.30, 99.90)	18.06
Park 2019	RFI/LR	87.00 (79.50, 92.00)	10.50
Wang 2019	DLRE	99.90 (98.00, 100.00)	18.91
Wang 2022	SVM/LR	60.70 (50.80, 70.00)	6.48
Zhang 2023	LASSO/LR	99.90 (99.80, 100.00)	19.31
Subtotal (I-squared = 95.7%, p = 0.000)		92.68 (89.68, 95.68)	100.00

NOTE: Weights are from random effects analysis

-100   0   100

Fig. 4 A forest graph showing the outcomes of meta-analysis of the sensitivity of radiomics in diagnosing fibrosis stages in training cohort

We have observed that some diagnostic accuracy indices were slightly higher in training cohorts than in validation cohorts. Wang et al. [33] who found that the accuracy of deep learning radiomics of elastography for the diagnosis of significant fibrosis was lower in the validation cohort in comparison with the training cohort suggested that this could be because of the lower heterogeneity in F0 and F1 groups and can be overcome possibly by adapting multiple strategies for fibrosis classification. Lu et al. [29] found that AUROC values increased for their deep learning radiomics of elastography when datasets had a higher prevalence of patients with F0 and F1 stages. They suggested that the inclusion of higher proportions of patients with F0 and F1 stages can yield better accuracy because about 80% of patients with chronic hepatitis B have F0 or F1 stage in the general population. In the present study, the pooled percentages of patients with F0 and F1 stages were 13% and 17% respectively.

Fibrosis develops by the excessive deposition of extracellular matrix in the liver as a response to wound healing after which satellite cells activate, higher levels of alpha smooth muscle actins are produced, and collagen I/II are synthesized. Such processes increase the stiffness of the liver progressively and may lead to cirrhosis. Fibrosis is usually associated with the accumulation of collagen fibers, not well-defined portal vein walls, and irregular hepatic vein margins. Gray-scale ultrasound images capture such information to reflect the scattering of fine structures. Moreover, coarse echotexture
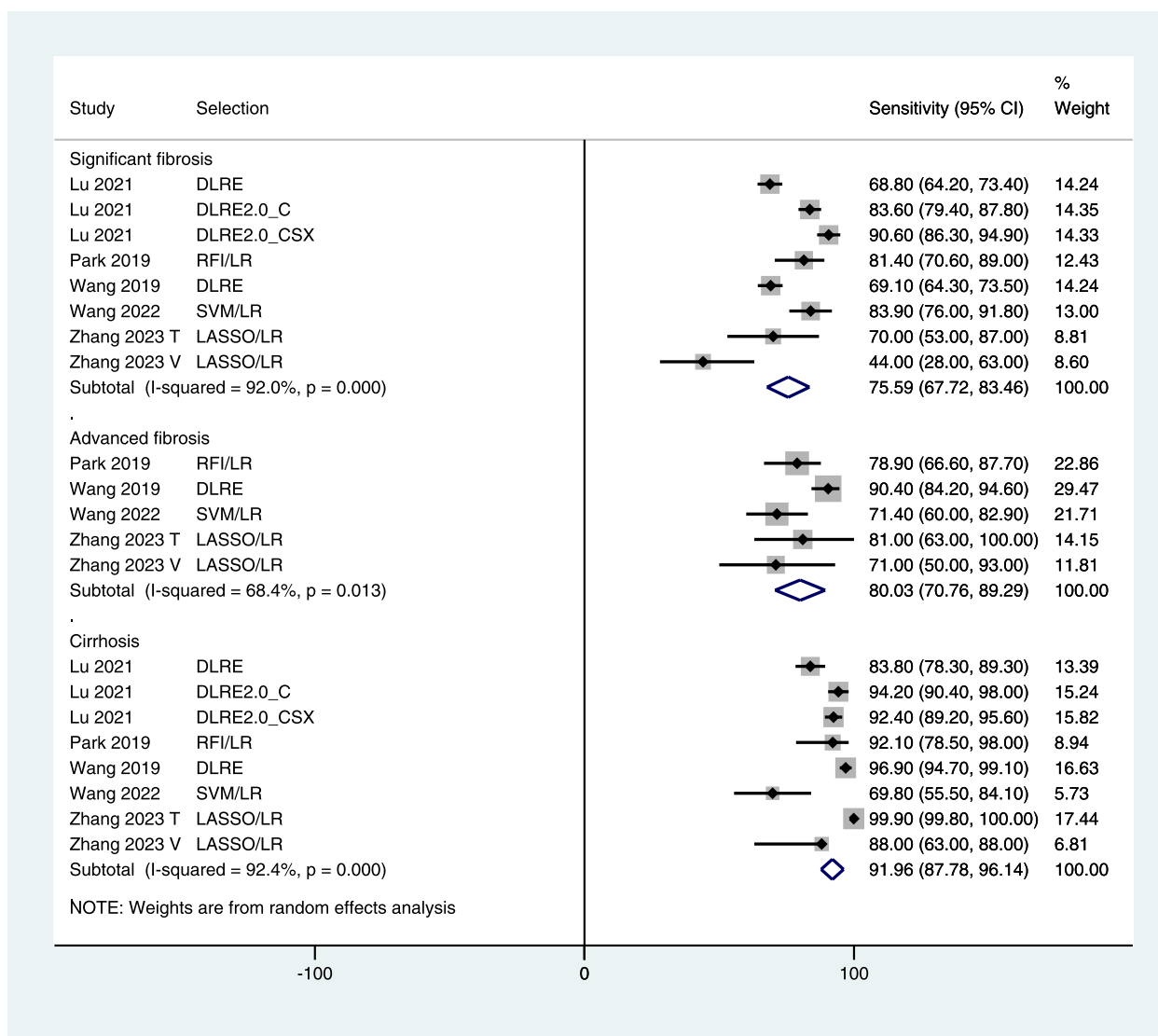
Study	Selection		Sensitivity (95% CI)	% Weight
Significant fibrosis				
Lu 2021	DLRE		68.80 (64.20, 73.40)	14.24
Lu 2021	DLRE2.0_C		83.60 (79.40, 87.80)	14.35
Lu 2021	DLRE2.0_CSX		90.60 (86.30, 94.90)	14.33
Park 2019	RFI/LR		81.40 (70.60, 89.00)	12.43
Wang 2019	DLRE		69.10 (64.30, 73.50)	14.24
Wang 2022	SVM/LR		83.90 (76.00, 91.80)	13.00
Zhang 2023 T	LASSO/LR		70.00 (53.00, 87.00)	8.81
Zhang 2023 V	LASSO/LR		44.00 (28.00, 63.00)	8.60
Subtotal (I-squared = 92.0%, p = 0.000)			75.59 (67.72, 83.46)	100.00
.				
Advanced fibrosis				
Park 2019	RFI/LR		78.90 (66.60, 87.70)	22.86
Wang 2019	DLRE		90.40 (84.20, 94.60)	29.47
Wang 2022	SVM/LR		71.40 (60.00, 82.90)	21.71
Zhang 2023 T	LASSO/LR		81.00 (63.00, 100.00)	14.15
Zhang 2023 V	LASSO/LR		71.00 (50.00, 93.00)	11.81
Subtotal (I-squared = 68.4%, p = 0.013)			80.03 (70.76, 89.29)	100.00
.				
Cirrhosis				
Lu 2021	DLRE		83.80 (78.30, 89.30)	13.39
Lu 2021	DLRE2.0_C		94.20 (90.40, 98.00)	15.24
Lu 2021	DLRE2.0_CSX		92.40 (89.20, 95.60)	15.82
Park 2019	RFI/LR		92.10 (78.50, 98.00)	8.94
Wang 2019	DLRE		96.90 (94.70, 99.10)	16.63
Wang 2022	SVM/LR		69.80 (55.50, 84.10)	5.73
Zhang 2023 T	LASSO/LR		99.90 (99.80, 100.00)	17.44
Zhang 2023 V	LASSO/LR		88.00 (63.00, 88.00)	6.81
Subtotal (I-squared = 92.4%, p = 0.000)			91.96 (87.78, 96.14)	100.00
NOTE: Weights are from random effects analysis				

-100   0   100

Fig. 5 A forest graph showing the outcomes of meta-analysis of the sensitivity of radiomics in diagnosing fibrosis stages in test/validation cohorts

and a mild increase in echogenicity of hepatic parenchyma are usually observed in cirrhosis [35].

So far, studies addressing radiomics lack robust validation in larger and clinically diverse settings which present reproducibility challenges. In the present review, we found that all except one study involved internal validation, and therefore, the synthesis of these outcomes remains inconclusive and dependent on future studies with larger sample sizes and better designs with special focus on external validation. Radiomics models without external validation are at increased risk of being specialized in specific radiographs that hamper generalizability due to overfitting [46].

Overfitting and multi-collinearity may affect radiomics models. During training, high-dimensional features may overfit and thus may yield optimistic outcomes. Moreover, traditional statistical models may not work adequately to deal with multicollinearity among textural features. To avoid this, it is suggested that the removal of unreliable or irrelevant features and the reduction of dimensions of predictors may yield better outcomes [30, 31]. The sensitivity of AdaBoost to noisy data or outliers makes it more suitable for cases facing overfitting problems. A frequently used classifier, the Support Vector Machine, uses preselected nonlinear mapping to map input parameters in a high-dimensional feature space to optimize feature classification. Random forest unifies

several weak predictor classifiers to make an accurate and stable predictor [28].

The AUROC is a performance metric to quantify the power of a model in discriminating cases from non-cases. An AUROC value can lie between 0 and 1. It combines the sensitivity and specificity of a marker/modality for the diagnosis of a precisely defined stage of fibrosis. Sensitivity is usually evaluated in patients with advanced fibrosis and specificity in non-advanced fibrosis [47]. However, the AUROC values can be biased if the fibrosis distribution in the study population differs from that of the whole population to which it is being applied [48]. Although biopsy is considered a gold standard for the diagnosis of fibrosis in the liver, it has a high rate of false positives and false negatives in comparison with the whole liver due to sampling error. An AUROC value of 0.82 for distinguishing between F2 and F1 when the entire liver was used as the reference index will inform approximately 20% error rate of the biopsy (false positive and false negative rates) compared with the entire liver. Thus, discordance in the staging of liver fibrosis between a modality such as radiomics and biopsy can be due to an error of the modality as well as due to an error of the biopsy [47].

Currently, the evidence regarding the role of radiomics in diagnosis of liver fibrosis and staging is constrained with several caveats. Quality of medical images acquired through different modalities may vary depending on several factors such as scanners, protocols, and personnel that can affect the reproducibility of radiomics output. A lack of standardization of image acquisition, preprocessing steps, extraction of features, and analyses also makes it difficult to compare radiomics outcomes of various studies performed under different settings. The etiology of fibrosis, progression, and the presence of comorbidities may also affect the accuracy of radiomics outcomes. Moreover, biological interpretation of radiomic features is lacking due to which it is difficult to associate radiomic features with histopathological characteristics.

Several limitations of the present study need to be considered while interpreting the outcomes of this review. An important limitation of the present study was the presence of high statistical heterogeneity in the meta-analyses. Although sources of heterogeneity could not be traced statistically, it is reasonable to assume that clinical and methodological heterogeneity might have played an influencing role. Authors utilized different methods for feature selection and classification, worked with a variety of software, and analyzed a highly variable number of features. Radiomics analyses were based on several radiological modalities and fewer studies were available to evaluate a particular modality in a pooled design. Most studies were retrospective in design due to which several types of biases could have been introduced. Inclusion and exclusion criteria differed substantially across the included studies that recruited several conditions of chronic liver disease including hepatitis B/C, autoimmune hepatitis, liver failure, early-stage cirrhosis, nonalcoholic fatty liver disease, and primary sclerosing cholangitis. Some studies could not be included because of the lack of variance data for diagnostic accuracy indices.

Conclusion

In this meta-analysis of 15 studies, the use of radiomics in staging liver fibrosis has been found to be associated with good diagnostic accuracy. However, the present-day outcome data are inconclusive regarding the use of radiomics in clinical practice owing to heterogeneity in methodology and outcomes of reviewed studies in which the radiomic evaluations were based on several radiological modalities subjected to a variety of analytical models yielding varying outcomes and lacking external validation. Non-invasiveness and the involvement of machine learning make radiomics an attractive option for decision support. The outcomes reported so far are promising and need to be validated in multi-center studies having larger datasets and better comparability and validation aspects in designs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-024-01272-x.

Supplementary Material 1.

Availability of data and materials
The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Cheemerla S, Balakrishnan M. Global epidemiology of chronic liver disease. Clin Liver Dis (Hoboken). 2021;17(5):365–70.
2. Centers for disease control and prevention. national center for health statistics. fast stats homepage. Chronic liver disease. Available at: https://www.cdc.gov/nchs/fastats/liver-disease.htm. Last accessed on March 16, 2024.
3. Li M, Wang ZQ, Zhang L, Zheng H, Liu DW, Zhou MG. Burden of cirrhosis and other chronic liver diseases caused by specific etiologies in China, 1990–2016: Findings from the Global Burden of Disease Study 2016. Biomed Environ Sci. 2020;33(1):1–10.
4. Zhai M, Liu Z, Long J, Zhou Q, Yang L, Zhou Q, et al. The incidence trends of liver cirrhosis caused by nonalcoholic steatohepatitis via the GBD study 2017. Sci Rep. 2021;11:5195. https://doi.org/10.1038/s41598-021-84577-z.
5. Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatch-aroenwitthaya P, et al. liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology. 2015;149(2):389–97.
6. Choi YS, Beltran TA, Calder SA, Padilla CR, Berry-Caban CS, Salyer KR. Prevalence of hepatic steatosis and fibrosis in the United States. Metab Syndr Relat Disord. 2022;20(3):141–7.
7. GBD 2017 Cirrhosis Collaborators. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol. 2020;5(3):245–66.
8. Chu LC, Park S, Kawamoto S, Yuille AL, Hruban RH, Fishman EK. Current status of radiomics and deep learning in liver imaging. J Comput Assist Tomogr. 2021;45:343–51.
9. Manning DS, Afdhal NH. Diagnosis and quantitation of fibrosis. Gastroenterology. 2008;134(6):1670–81.
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77.
11. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749–62. https://doi.org/10.1038/nrclinonc.2017.141.
12. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. Radiol Med. 2021;126(10):1296–311.
13. Lan GY, Guo Y, Zhang XY, Cai XL, Shi Y. Value of radiomic analysis of data from magnetic resonance elastography for diagnosing fibrosis stages in patients with hepatitis B/C. Chin J Acad Radiol. 2019;1:74–84.
14. Park HJ, Park B, Lee SS. Radiomics and deep learning: hepatic applications. Korean J Radiol. 2020;21:387–401.
15. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges. Theranostics. 2019;9(5):1303–22.
16. Zhang YP, Zhang XY, Cheng YT, Li B, Teng XZ, Zhang J, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. Mil Med Res. 2023;10(1):22. https://doi.org/10.1186/s40779-023-00458-8.
17. Neisius U, El-Rewaidy H, Nakamori S, Rodriguez J, Manning WJ, Nezafat R. Radiomic analysis of myocardial native T1 imaging discriminates between hypertensive heart disease and hypertrophic cardiomyopathy. JACC Cardiovasc Imaging. 2019;12(10):1946–54.
18. Neisius U, El-Rewaidy H, Kucukseymen S, Tsao CW, Mancio J, Nakamori S, et al. Texture signatures of native myocardial T1 as novel imaging markers for identification of hypertrophic cardiomyopathy patients without scar. J Magn Reson Imaging. 2020;52(3):906–19.
19. Cuadra MB, Favre J, Omoumi P. Quantification in musculoskeletal imaging using computational analysis and machine learning: Segmentation and radiomics. Semin Musculoskelet Radiol. 2020;24(1):50–64. https://doi.org/10.1055/s-0039-3400268.
20. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. Skeletal Radiol. 2020;49(2):183–97.
21. Park YW, Choi D, Lee J, Ahn SS, Lee SK, Lee SH, et al. Differentiating patients with schizophrenia from healthy controls by hippocampal subfields using radiomics. Schizophr Res. 2020;223:337–44.
22. Park YW, Choi YS, Kim SE, Choi D, Han K, Kim H, et al. Radiomics features of hippocampal regions in magnetic resonance imaging can differentiate medial temporal lobe epilepsy patients from healthy controls. Sci Rep. 2020;10:19567.
23. Bang M, Park YW, Eom J, Ahn SS, Kim J, Lee SK, et al. An interpretable radiomics model for the diagnosis of panic disorder with or without agoraphobia using magnetic resonance imaging. J Affect Disord. 2022;305:47–54.
24. Zhao K, Ding Y, Han Y, Fan Y, Alexander-Bloch AF, Han T, et al. Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer's disease: diagnosis, longitudinal progress and biological basis. Sci Bull. 2020;65:1103–13.
25. Liu F, Ning Z, Liu Y, Liu D, Tian J, Luo H, et al. Development and validation of a radiomics signature for clinically significant portal hypertension in cirrhosis (CHESS1701): a prospective multicenter study. EBioMedicine. 2018;36:151–8.
26. Chen ZW, Tang K, Zhao YF, Chen YZ, Tang LJ, Li G, et al. Radiomics based on fluoro-deoxyglucose positron emission tomography predicts liver fibrosis in biopsy-proven MAFLD: a pilot study. Int J Med Sci. 2021;18:3624–30.
27. Hu P, Chen L, Zhong Y, Lin Y, Yu X, Hu X, et al. Effects of slice thickness on CT radiomics features and models for staging liver fibrosis caused by chronic liver disease. Japanese J Radiol. 2022;40:1061–8.
28. Li W, Huang Y, Zhuang BW, Liu GJ, Hu HT, Li X, et al. Multiparametric ultrasomics of significant liver fibrosis: A machine learning-based analysis. Eur Radiol. 2019;29:1496–506.
29. Lu X, Zhou H, Wang K, Jin J, Meng F, Mu X, et al. Comparing radiomics models with different inputs for accurate diagnosis of significant fibrosis in chronic liver disease. Eur Radiol. 2021;31:8743–54.
30. Park HJ, Lee SS, Park B, Yun J, Sung YS, Shim WH, et al. Radiomics analysis of Gadoxetic Acid–enhanced MRI for staging liver fibrosis. Radiology. 2019;290:380–7.
31. Qiu QT, Zhang J, Duan JH, Wu SZ, Ding JL, Yin Y. Development and validation of radiomics model built by incorporating machine learning for identifying liver fibrosis and early-stage cirrhosis. Chin Med J. 2020;133(22):2653–9.
32. Sim KC, Kim MJ, Cho Y, Kim HJ, Park BJ, Sung DJ, et al. Diagnostic feasibility of magnetic resonance elastography radiomics analysis for the assessment of hepatic fibrosis in patients with nonalcoholic fatty liver disease. J Comput Assist Tomogr. 2022;46:505–13.
33. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut. 2019;68:729–41.
34. Wang J, Tang S, Mao Y, Wu J, Xu S, Yue Q, et al. Radiomics analysis of contrast-enhanced CT for staging liver fibrosis: an update for image biomarker. Hepatol Int. 2022;16:627–39.
35. Xue LY, Jiang ZY, Fu TT, Wang QM, Zhu YL, Dai M, et al. Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis. Eur Radiol. 2020;30:2973–83.
36. Yin Y, Yakar D, Dierckx RAJO, Mouridsen KB, Kwee TC, de Haas RJ. Combining hepatic and splenic CT radiomic features improves radiomic analysis performance for liver fibrosis staging. Diagnostics. 2022;12:550. https://doi.org/10.3390/diagnostics12020550.
37. Zhang D, Cao Y, Sun Y, Xia Zhao X, Peng C, Zhao J, et al. Radiomics nomograms based on R2* mapping and clinical biomarkers for staging of liver fibrosis in patients with chronic hepatitis B: a single-center retrospective study. Eur Radiol. 2023;33:1653–67.
38. Zhao R, Zhao H, Ge YQ, Zhou FF, Wang LS, Yu HZ, et al. Usefulness of noncontrast MRI-based radiomics combined clinic biomarkers in stratification of liver fibrosis. Canadian J Gastroenterol Hepatol. 2022;2022:2249447. Article ID 2249447.
39. Zheng R, Shi C, Wang C, Shi N, Qiu T, Chen W, et al. Imaging-based staging of hepatic fibrosis in patients with hepatitis B: A dynamic radiomics model based on Gd-EOB-DTPA-enhanced MRI. Biomolecules. 2021;11:307.

40. Cui E, Long W, Wu J, Li Q, Ma C, Lei Y, et al. Predicting the stages of liver fibrosis with multiphase CT radiomics based on volumetric features. Abdom Radiol. 2021;46:3866–76.
41 Duan YY, Qin J, Qiu WQ, Li SY, Li C, Liu AS, et al. Performance of a generative adversarial network using ultrasound images to stage liver fibrosis and predict cirrhosis based on a deep-learning radiomics nomogram. Clin Radiol. 2022;77(10):e723–31.
42. Zhou Z, Zhang Z, Gao A, Tai DI, Wu S, Tsui PH. Liver fibrosis assessment using radiomics of ultrasound homodyned-K imaging based on the artificial neural network estimator. Ultrasonic Imaging. 2022;44(5–6):229–41.
43. Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S. Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid–enhanced hepatobiliary phase MR images. Radiology. 2018;287:146–55.
44. Choi KJ, Jang JK, Lee SS, Sung YS, Shim WH, Kim HS, et al. Development and validation of a deep learning system for staging liver fibrosis by using contrast agent–enhanced CT images in the liver. Radiology. 2018;289:688–97.
45. Zheng W, Guo W, Xiong M, Chen X, Gao L, Song Y, et al. Clinic-radiological features and radiomics signatures based on Gd-BOPTA-enhanced MRI for predicting advanced liver fibrosis. European Radiology. 2023;33:633–44.
46. Yamada A, Kamagata K, Hirata K, Ito R, Nakaura T, Ueda D, et al. Clinical applications of artificial intelligence in liver imaging. Radiol Med. 2023;128(6):655–67.
47. Poynard T, Halfon P, Castera L, Charlotte F, Le Bail B, Munteanu M, et al. Variability of the area under the receiver operating characteristic curves in the diagnostic evaluation of liver fibrosis markers: impact of biopsy length and fragmentation. Aliment Pharmacol Ther. 2007;25(6):733–9.
48. Guha IN, Myers RP, Patel K, Talwalkar JA. Biomarkers of liver fibrosis: What lies beneath the receiver operating characteristic curve? Hepatology. 2011;54:1454–62.

Publisher's Note