

An ultrasound image-based deep multi-scale texture network for liver fibrosis grading in patients with chronic HBV infection

Dongsheng Ruan^{1,2}  | Yu Shi²  | Linfeng Jin² | Qiao Yang³ | Wenwen Yu⁴ | Haotang Ren² | Weiyang Zheng²  | Yongping Chen⁵ | Nenggan Zheng¹ | Min Zheng² 

¹Qiushi Academy for Advanced Studies, Zhejiang University, Hangzhou, Zhejiang, P. R. China

²State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, P. R. China

³Department of Infectious Diseases, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, P. R. China

⁴Department of Infectious Diseases, Beilun People's Hospital, Ningbo, P. R. China

⁵Wenzhou Key Laboratory of Hepatology, Department of Infectious Diseases, Hepatology Institute of Wenzhou Medical University, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, P. R. China

Correspondence

Nenggan Zheng, Qiushi Academy for Advanced Studies, Zhejiang University, Hangzhou, Zhejiang, P. R. China.
Email: zng@cs.zju.edu.cn

Min Zheng, State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, P. R. China.
Email: minzheng@zju.edu.cn

Funding information

13-5 State S&T Projects of China, Grant/Award Number: 2018ZX1030206; Natural Science Foundation of China, Grant/Award Number: 81871646 and 81870425

Handling Editor: Junko Tanaka

Abstract

Background & Aims: The evaluation of the stage of liver fibrosis is essential in patients with chronic liver disease. However, due to the low quality of ultrasound images, the non-invasive diagnosis of liver fibrosis based on ultrasound images is still an outstanding question. This study aimed to investigate the diagnostic accuracy of a deep learning-based method in ultrasound images for liver fibrosis staging in multicentre patients.

Methods: In this study, we proposed a novel deep learning-based approach, named multi-scale texture network (MSTNet), to assess liver fibrosis, which extracted multi-scale texture features from constructed image pyramid patches. Its diagnostic accuracy was investigated by comparing it with APRI, FIB-4, Forns and sonographers. Data of 508 patients who underwent liver biopsy were included from 4 hospitals. The area-under-the ROC curve (AUC) was determined by receiver operating characteristics (ROC) curves for significant fibrosis (\geq F2) and cirrhosis (F4).

Results: The AUCs (95% confidence interval) of MSTNet were 0.92 (0.87-0.96) for \geq F2 and 0.89 (0.83-0.95) for F4 on the validation group, which significantly outperformed APRI, FIB-4 and Forns. The sensitivity and specificity of MSTNet (85.1% (74.5%-92.0%) and 87.6% (78.0%-93.6%)) were better than those of three sonographers in assessing \geq F2.

Conclusions: The proposed MSTNet is a promising ultrasound image-based method for the non-invasive grading of liver fibrosis in patients with chronic HBV infection.

Dongsheng Ruan, Yu Shi, and Linfeng Jin contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Liver International* published by John Wiley & Sons Ltd.

KEYWORDS

deep learning, image classification, liver fibrosis, ultrasound images

1 | INTRODUCTION

Chronic hepatitis B virus (HBV) infection remains a major global health burden, with approximately 250 million individuals infected worldwide and 800 000 deaths annually.¹⁻⁴ Liver fibrosis is the result of structure and function alteration in chronic liver diseases. Mounting evidences have demonstrated that the severity of liver fibrosis correlates with the risk of long-term complications, such as decompensated cirrhosis and hepatocellular carcinoma, and is one of the significant determinants of the prognosis of patients with chronic HBV infection.⁵⁻⁷ In this regard, patients with significant fibrosis, but the alanine transaminase (ALT) levels are normal or only minimally elevated, are recommended to receive antiviral therapy to prevent ongoing progression of fibrosis.⁸⁻¹⁰ Therefore, correct grading of liver fibrosis is of utmost importance for assessing disease severity and determining prognosis and clinical intervention in patients with chronic HBV infection.

Liver biopsy is the gold-standard criteria for the assessment of liver fibrosis but is an invasive procedure that limits the use for widespread screening and carries a risk of rare but potentially life-threatening complications.¹¹⁻¹³ Besides, liver biopsy has technical limitations such as sampling error and interobserver variabilities.¹⁴⁻¹⁶ To overcome these limitations, non-invasive tests based on a panel of serum biomarkers, such as aspartate transaminase-to-platelet ratio index (APRI), fibrosis index based on four factors (FIB-4) and Forns test, are first developed to assess liver fibrosis.¹⁶⁻¹⁹ However, most of these serum biomarkers are not liver-specific, and various individual parameters influence their levels. In recent years, liver stiffness measurement (LSM) technologies, including transient elastography (TE) and shear wave elastography (SWE), have been applied for the diagnosis of liver fibrosis.²⁰⁻²² LSM is a rapid, convenient and liver-specific test but still has limitations. The major flaw is that the stiffness values can be affected by several clinical settings such as obesity, ascites, alanine aminotransferase (ALT) flares, extra-hepatic cholestasis and so on.

23-25

Recently, the B-mode ultrasound imaging-based diagnostic tests have attracted much attention since B-mode ultrasound images are more informative than 1-dimensional ultrasound. Though several computer-aided diagnostic techniques have been proposed to assess liver fibrosis, they are limited by insufficient texture extraction capability, handcrafted feature statistics and low diagnostic accuracy.²⁶⁻³³ Here, we introduced a deep learning-based diagnostic method, named multi-scale texture network (MSTNet) that can use a pure pyramid residual convolution structure to capture multi-scale texture features of ultrasound images. In this way, the coarse-and-fine texture features can be effectively extracted from ultrasound

Key points

- Liver fibrosis is prevalent in patients with chronic HBV infection.
- Liver biopsy is still an invasive diagnosis that carries a risk of complications.
- The performance of existing non-invasive methods still needs to be improved.
- The proposed deep-learning ultrasound image-based method shows a good potential for the non-invasive grading liver fibrosis in patients with chronic HBV infection.

images. In the present retrospective multicentre study, we evaluated the accuracy of MSTNet in staging liver fibrosis compared with three serum biomarker-based tests and sonographers.

2 | PATIENTS AND METHODS

2.1 | Patients

Patients with chronic HBV infection who underwent B-ultrasound test and liver biopsy were enrolled in First Affiliated Hospital of Zhejiang University, Sir Run Run Hospital of Zhejiang University, First Affiliated Hospital of Wenzhou Medical University and Beilun People's Hospital between January 2014 and September 2019. The diagnosis of chronic HBV infection is defined as the persistence of serum HBsAg positivity for over 6 months. All enrolled patients had not been treated with nucleoside analogue. Patients were excluded: (i) with the presence of significant alcohol consumption (30 g/day for males; 20 g/day for females), other viral hepatitis, autoimmune hepatitis, primary biliary cholangitis, primary sclerosing cholangitis, Wilson disease or other hereditary liver diseases; (ii) liver transplant; (iii) HIV infection; (iv) with low-quality ultrasound image. The study complied with the principles of the Declaration of Helsinki and was approved by the ethics committee of each participating hospital. Written consent was provided by each participant or their legal representatives.

2.2 | Data collection and model calculation

For each patient, the following information was collected and input into a pre-specified, electronic data collection form: demographic data (age and sex), laboratory data [alanine aminotransferase (ALT),

aspartate aminotransferase (AST), platelet count, γ -glutamyl transpeptidase (GGT), globulin, serum bilirubin, total cholesterol (TCH), HBeAg, liver B-ultrasound images and histological diagnosis of liver biopsy. All the liver B-ultrasound images were reviewed and re-diagnosed by three independent experienced sonographers.

Three serum biomarker-based tests for liver fibrosis were calculated as follows: APRI: $\frac{AST/(ULN)}{PLT} \times 100$, FIB-4: $\frac{age \times AST}{PLT \times \sqrt{ALT}}$ and Forns: $7.811 - 3.131 \times \ln PLT + 0.781 \times \ln GGT + 3.467 \times \ln age - 0.014 \times TCH$, where ULN is the value for the AST upper limit of normal and is set as 40 IU/L in our study.

2.3 | Liver biopsy and histology evaluation

Liver biopsies were routinely stained with haematoxylin and eosin. All the liver biopsies were reviewed by a senior pathologist in each participating centre. To reduce the sampling errors of needle biopsy, the specimens of liver biopsy were at least 20 mm length and 11 portal tracts. A 16 or 18-gauge needle was used in a liver lobule to avoid transecting the portal tracts. METAVIR scoring system was adopted for the histologic staging of fibrosis. The grades of $\geq F2$ and F4 denote significant fibrosis and cirrhosis, respectively.

2.4 | MSTNet

To fully capture liver fibrosis texture of ultrasound images, we introduce the deep learning-based diagnostic approach MSTNet, shown in Figure 1 and Table 1 that consists of image pyramid patches and distribute-gather attention network (DGANet).

2.4.1 | Image pyramid patches

The purpose of image pyramid patches was to enable DGANet to learn different scales of liver fibrosis textures and make full use of the texture information in the liver parenchyma. Here, we cropped three image patches of different resolutions from the liver parenchyma in the original ultrasound image: 60×60 , 100×100 and 120×120 . One to four image patches of each resolution were

cropped to ensure the diversity of texture, and the number of the cropped image patches depended on the resolution of the liver parenchyma area. The overlap rate of any two image patches was $<50\%$. Figure 2 shows an example of building image pyramid patches. It is worth noting that blood vessels, haemangiomas and cysts in the image patch should be avoided as much as possible. Considering the similarity of the liver parenchyma texture, we assigned the liver fibrosis grading of the entire ultrasound image to each image patch. In this way, the training data was augmented, thereby effectively alleviating the network overfitting.

2.4.2 | DGANet

DGANet mainly consists of DGA blocks, Figure 3A, to adaptively capture multi-scale texture features and transition layers to down-sample the feature maps of two adjacent stages. DGA block has two units: DG unit and multi-branch attention unit.

TABLE 1 Architecture of MSTNet

Layer name	50-layer
Convolution	7×7 conv, stride 1
Pooling	3×3 max pool, stride 2
Stage 1	$\left[\begin{array}{c} 1 \times 1, 64 \\ [(3 \times 3, D=1), (3 \times 3, D=2)], 64 \\ 1 \times 1, 256 \end{array} \right] \times 3$
Transition 1	2×2 average pool, stride 2
Stage 2	$\left[\begin{array}{c} 1 \times 1, 256 \\ [(3 \times 3, D=1), (3 \times 3, D=2)], 256 \\ 1 \times 1, 512 \end{array} \right] \times 4$
Transition 2	2×2 average pool, stride 2
Stage 3	$\left[\begin{array}{c} 1 \times 1, 512 \\ [(3 \times 3, D=1), (3 \times 3, D=2)], 512 \\ 1 \times 1, 1024 \end{array} \right] \times 6$
Transition 3	2×2 average pool, stride 2
Stage 4	$\left[\begin{array}{c} 1 \times 1, 512 \\ [(3 \times 3, D=1), (3 \times 3, D=2)], 512 \\ 1 \times 1, 2048 \end{array} \right] \times 3$
Classification	Average pool, 3-d FC, softmax

Note: $(3 \times 3, D = *)$ denotes 3×3 grouped convolution with dilation rate *. The group G is set to 16.

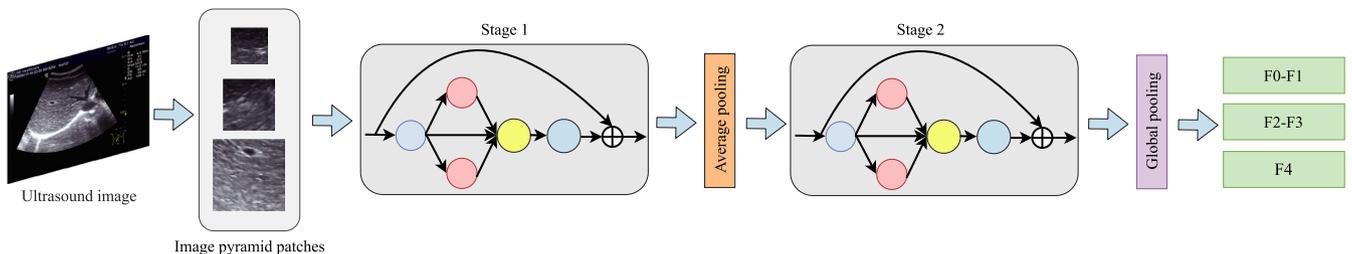
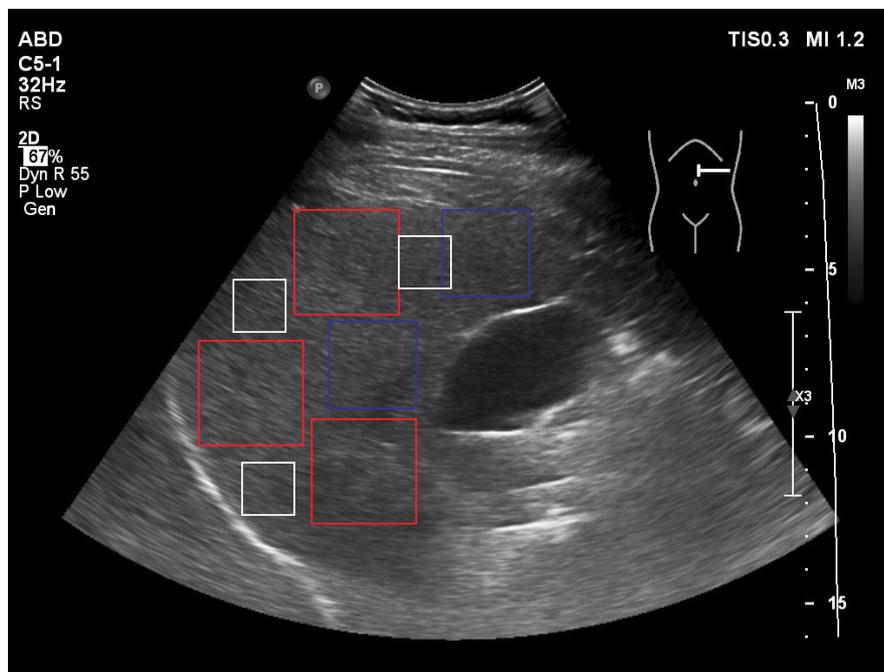


FIGURE 1 Schematic illustration of MSTNet. MSTNet consists of image pyramid patches cropped from ultrasound images and distribute-gather attention network. It has multiple stages, and each stage has multiple distribute-gather attention blocks. The transition layer is introduced to reduce the size of the feature maps between two continuous stages

FIGURE 2 An example of image pyramid patches. Red, blue and white-bounding boxes represent image patches of 120×120 pixels, 100×100 pixels and 60×60 pixels, respectively



As shown in Figure 3B, the DG unit comprised a squeeze layer, a distribute-gather (DG) layer and an excitation layer. The squeeze layer had a 1×1 convolution layer that was mainly responsible for reducing dimensions. The DG layer had three branches, which was used to capture multi-scale texture features of ultrasound images. Specifically, branch 1 was an identity mapping to reuse the features of the last layer. For branches 2 and 3, 3×3 grouped convolutions with dilation rates 1 and 2 were used to perform group-wise spatial convolutions. The excitation layer was also a 1×1 convolution filter that aims to fuse hierarchical texture features.

The multi-branch attention unit, depicted in Figure 4, aimed to adaptively select the scale texture information to best match the grading of liver fibrosis, which first applied 1×1 grouped convolutions to fuse the features of each branch across channels. Then, a global average pooling operator was used to produce global contexts of these spatial features across spatial dimensions. Further, two fully connected layers followed by a *softmax* normalizer were applied to model the relationships between branches to obtain the attention vector that represented the importance of each texture scale. Finally, the final feature maps were rescaled via the attention vector.

The transition layer had a 2×2 average pooling operation to reduce the size of feature maps between two continuous stages. The first residual block in each stage was responsible for increasing the channel dimension.

2.5 | Overall framework

Figure 1 shows the overall architecture of the network. First, the image pyramid patches cropped from the ultrasound images were input to DGANet to extract multi-scale texture features by

multiple DGA blocks and transition layers. Then, multi-scale texture features were squeezed to a texture representation vector by global average pooling. Finally, the representation vector was transformed to the score of each fibrosis stage via a fully connected layer.

2.6 | Statistical analysis

To ensure data balance, 70% of patients from each hospital were randomly selected as the training group and the remaining patients as the validation group. Continuous variables were expressed as mean \pm standard deviation and compared by Student's *t* test or Mann-Whitney test. Nominal variables were expressed as a number (percent) and examined by χ^2 test. Area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, positive and negative predictive values (PPV and NPV) and positive and negative diagnostic likelihood ratios (LR+ and LR-) with the corresponding 95% confidence intervals (95% CI) were estimated to evaluate the diagnostic accuracy of different approaches according to Yoden Index. All the statistical analyses were performed using python for Ubuntu (V.3.6.5) and MedCalc software (V19.0.7) for Windows. *P* values below .05 were considered significant.

3 | RESULTS

3.1 | Training framework

As shown in Figure 5, the whole framework of our approach was summarized as three phases. Phase 1 was pre-training that trained

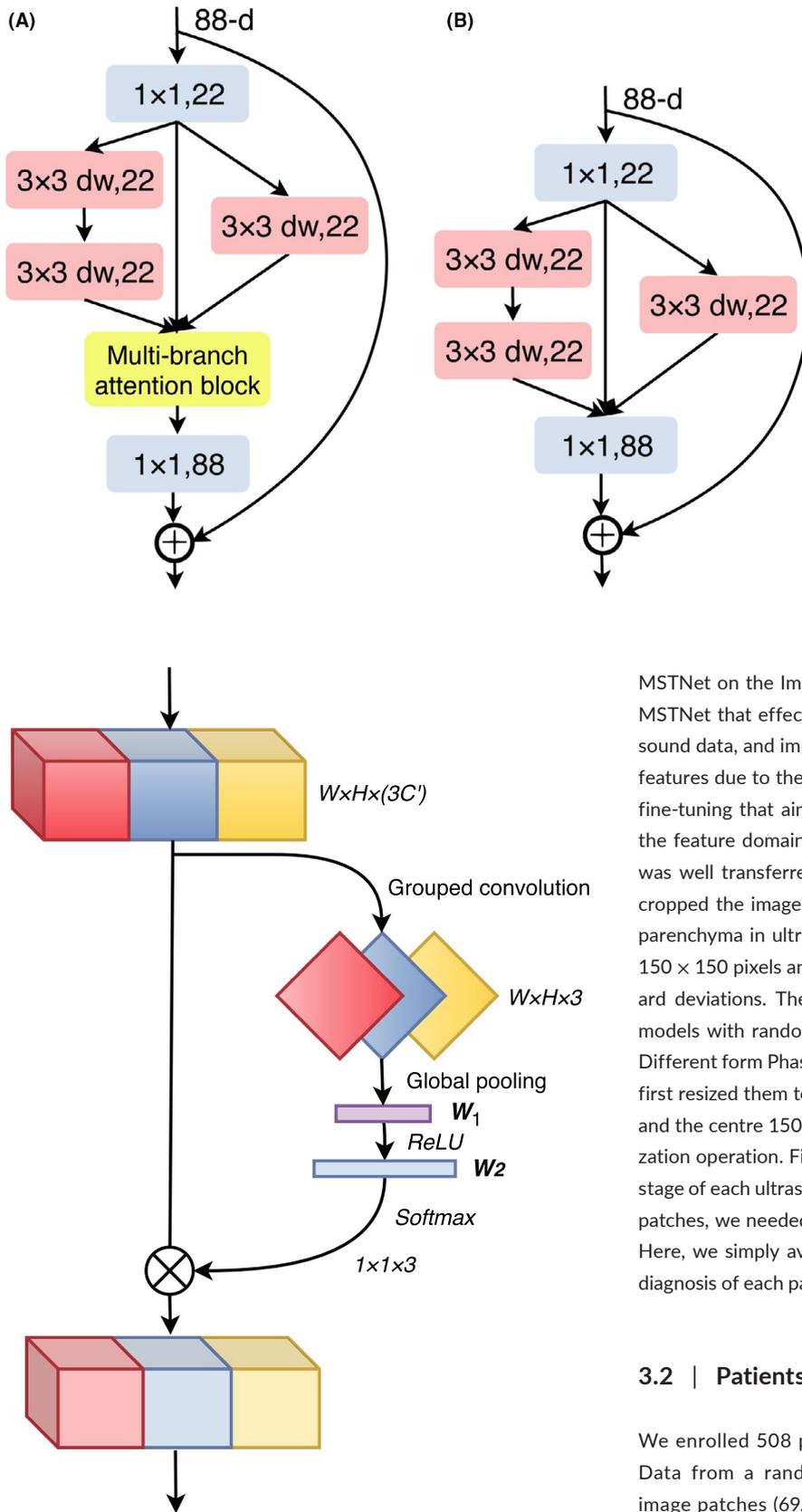


FIGURE 3 Schematic illustrations of (A) DGA unit and (B) DG block. DGA block consists of a DG unit and a multi-branch attention unit. G and D denote group number and dilation rate of grouped convolutional filter, respectively

FIGURE 4 A multi-branch attention unit. W, H and C' denote the spatial width, height and channel dimension of the feature map. W_1 and W_2 denote the weights of the FC layers. ReLU denotes a rectified linear unit that is an activation function

MSTNet on the ImageNet dataset. It enabled proper initialization of MSTNet that effectively alleviated the problem of insufficient ultrasound data, and improved the ability of MSTNet to capture low-level features due to the transferability of low-level features. Phase 2 was fine-tuning that aimed to adapt the feature domain of ImageNet to the feature domain of ultrasonic liver fibrosis data, so that MSTNet was well transferred to the ultrasonic domain. Specifically, we first cropped the image patches with different resolutions from the liver parenchyma in ultrasound images. Then, all patches were resized to 150×150 pixels and normalized using the channel means and standard deviations. The normalized image patches were input into the models with random horizontal flip. Phase 3 was the testing phase. Different from Phase 2, only 150×150 image patches were used. We first resized them to 170×170 pixels and then took the four corners and the centre 150×150 crop as input to MSTNet with the normalization operation. Finally, the output of MSTNet was the liver fibrosis stage of each ultrasound image. Since each patient had multiple image patches, we needed to synthesize the diagnosis of all image patches. Here, we simply averaged the results of all the patches as the final diagnosis of each patient.

3.2 | Patients' characteristics

We enrolled 508 patients from 4 Chinese hospitals in this study. Data from a randomly split-sample of 353 patients with 5690 image patches (69.5%) were used to train the model MSTNet and the remaining 155 patients with 1414 image patches to validate the model. The detailed characteristics are shown in Table 2. There were no significant differences between the training and the validation groups in any of the assessed variables ($P > .05$).

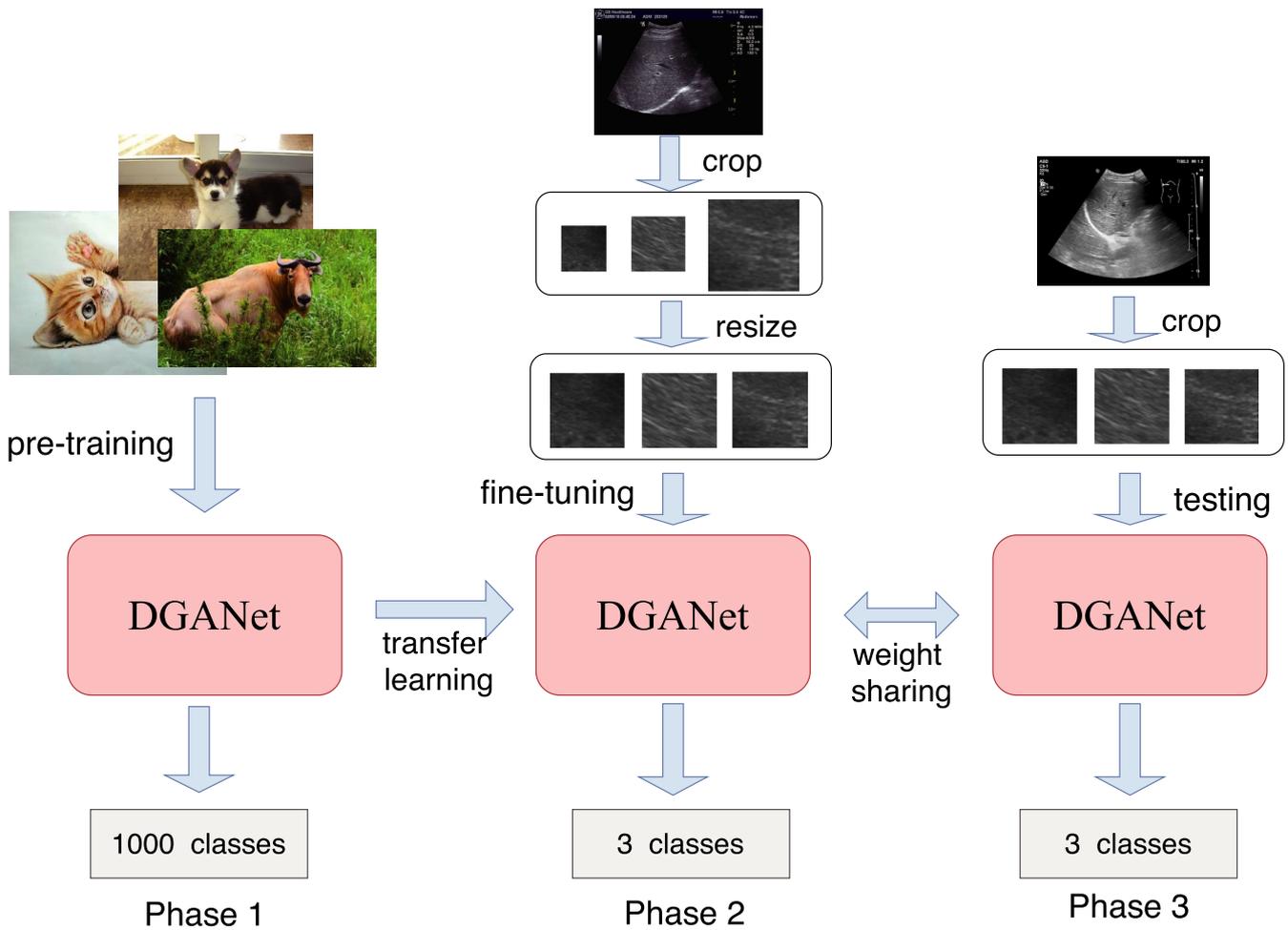


FIGURE 5 Overall framework of liver fibrosis grading. Phase 1: Pre-train DGANet on the ImageNet dataset. Phase 2: Fine-tune DGANet in the training group. Phase 3: Test in the validation group

Fifty-two percent of the patients were classified as normal or mild fibrosis (F0-F1), 27.5% as moderate fibrosis (F2-F3) and 20.5% as cirrhosis (F4).

3.3 | Diagnostic accuracy of MSTNet in comparison with APRI, FIB-4 and Forns

First, we evaluated MSTNet in comparison with three biomarker-based scores APRI, FIB-4 and Forns. Figure 6 and Table 3 show the ROC curves and diagnostic performance between MSTNet and three scores for the assessment of liver fibrosis stages in training and validation groups, respectively. It was clear that MSTNet achieved the best diagnostic performance compared with other methods for classifying $\geq F2$ and F4 in both groups. In the training group, the AUCs of MSTNet reached 1.00 for $\geq F2$ and 1.00 for F4, which were significantly higher than other methods ($P < .0001$, Table 3). Compared with Forns that performed second better, MSTNet obtained 0.20 and 0.11 AUC gains for classifying $\geq F2$ and F4, respectively (Figure 6A,B). The sensitivity, specificity and other

statistics also showed that MSTNet was superior to the other three approaches (Table 3).

In the validation group, MSTNet reached 0.92 and 0.89 AUCs for the diagnosis of $\geq F2$ and F4, which were slightly inferior to that in the training group but remained better than the other three scores ($P < .001$ or $P < .0001$, Table 3). In particular, the AUC of MSTNet increased 0.21, 0.25 and 0.15 in comparison to FIB-4, APRI and Forns, respectively, when judging advanced fibrosis ($\geq F2$). Regarding sensitivity and specificity, MSTNet outperformed FIB-4 and APRI significantly. Compared with Forns, MSTNet had better sensitivity and specificity for classifying $\geq F2$ (sensitivity: 85.1% vs 75.0%, specificity: 87.6% vs 69.5%). For F4, the sensitivity of MSTNet was slightly lower (87.9% vs 91.0%), but the specificity was significantly higher (78.1% vs 66.6%).

3.4 | Diagnostic accuracy versus resolutions of the image patches

We next investigated the effect of the input image patch resolution on the diagnostic accuracy of MSTNet. We compared four different

TABLE 2 Baseline characteristics of patients in the training and the validation groups

Variables	All patients	Training group	Validation group	P values
Number (%)	508	353 (70%)	155 (30%)	—
Age (years)	41.6 ± 13.0	41.4 ± 13.1	42.0 ± 12.7	0.65
Gender (male)	303 (70.1%)	211 (70.3%)	92 (69.7%)	0.68
PLT (10 ⁹ /L)	172.4 ± 61.5	169.3 ± 62.7	179.3 ± 58.0	0.12
ALT (IU/L)	60.0 ± 108.2	61.6 ± 115.1	56.3 ± 90.5	0.64
AST (IU/L)	43.1 ± 61.4	43.4 ± 61.5	42.3 ± 61.0	0.87
GGT (IU/L)	40.0 ± 73.1	34.6 ± 46.2	52.2 ± 111.0	0.44
GLOB (10 ⁹ /L)	29.3 ± 7.2	28.9 ± 6.8	30.1 ± 7.9	0.11
TB (μmol/L)	17.4 ± 18.7	16.2 ± 10.1	20.1 ± 29.9	0.41
TCH (mmol/L)	3.9 ± 0.9	4.0 ± 0.9	3.9 ± 0.9	0.55
HBeAg (±)	184/140	129/96	55/44	0.93
HBV DNA	5.6 × 10 ⁷ ± 1.9 × 10 ⁸	5.0 × 10 ⁷ ± 1.8 × 10 ⁸	6.9 × 10 ⁷ ± 2.1 × 10 ⁸	0.39
Fibrosis stages				
F0-F1	264 (52.0%)	184 (52.1%)	80 (51.6%)	—
F2-F3	140 (27.5%)	97 (27.5%)	43 (27.7%)	—
F4	104 (20.5%)	72 (20.4%)	32 (20.7%)	0.99
Inflammation grades				
A0	5 (1.2%)	4 (1.4%)	1 (0.8%)	—
A1	229 (54.0%)	164 (55.4%)	65 (50.8%)	—
A2	147 (34.7%)	97 (32.7%)	50 (39.0%)	—
A3	43 (10.1%)	37 (10.5%)	12 (9.4%)	0.63

Note: Statistics of continuous variables are reported as mean standard deviation (*P* value according to Student's *t* test or Mann-Whitney test). Categorical variables are reported as frequencies (*P* value according to χ^2 test).

P values were calculated between the training and validation groups. There were no significant differences between the training and the validation groups in any of the variables (*P* > .05).

Abbreviations: ALT, alanine amino transferase; AST, aspartate amino transferase; GGT, gamma glutamyl transpeptidase; GLOB, globulin; HBeAg, hepatitis Be antigen; PLT, platelet counts; TB, total bilirubin; TCH, total cholesterol.

patch resolutions: 60 × 60, 100 × 100, 120 × 120 and 150 × 150. Table 4 and Figure 7 show the diagnostic accuracy based on the resolutions of the image patches. The AUC of MSTNet was lower when the resolution of the image patch was smaller (60 × 60) and can be significantly improved as the patch resolution increased to 120 × 120 (≥F2: 0.84 → 0.90; F4: 0.84 → 0.88). But the improvement in diagnostic accuracy was not significant when the patch resolution increased to 150 × 150 from 120 × 120 (≥F2, AUC: T, ~0.99, V, ~0.92; F4, AUC: T, ~0.99, V, ~0.88).

3.5 | Diagnostic accuracy of MSTNet in comparison with sonographers

To further validate the advantage of MSTNet, we compared our approach with three sonographers. Each sonographer independently judged advanced liver fibrosis according to the ultrasound images (≥F2 in histology as the reference). Table 5 compared the diagnostic accuracy of MSTNet and that of each sonographer. We found that MSTNet performed better than the sonographers. Specifically,

the sensitivity of MSTNet was comparable to that of three sonographers (85.1% vs 80.6%, 85.4% and 82.2%), while the specificity was significantly higher (87.6% vs 65.2%, 58.7% and 78.3%), the other diagnostic parameters such as negative and positive predictive value were also higher in MSTNet approach.

3.6 | Diagnostic performance of MSTNet in patients with and without fatty liver

In order to examine the influence of fatty liver on the diagnostic performance of MSTNet, we compared the MSTNet performance in patients with and without fatty liver. Specifically, cases with fatty liver were selected from 508 patients (93 patients). The non-fatty liver patients were randomly selected from the remaining patients at 1:1 ratio. Considering the insufficient number of patients with liver cirrhosis in the fatty liver group, we only evaluated the diagnostic performance of MSTNet for diagnosing significant fibrosis.

Figure 8 showed the ROC curves of MSTNet in patients with and without fatty liver. The AUC was similar in both fatty liver group and

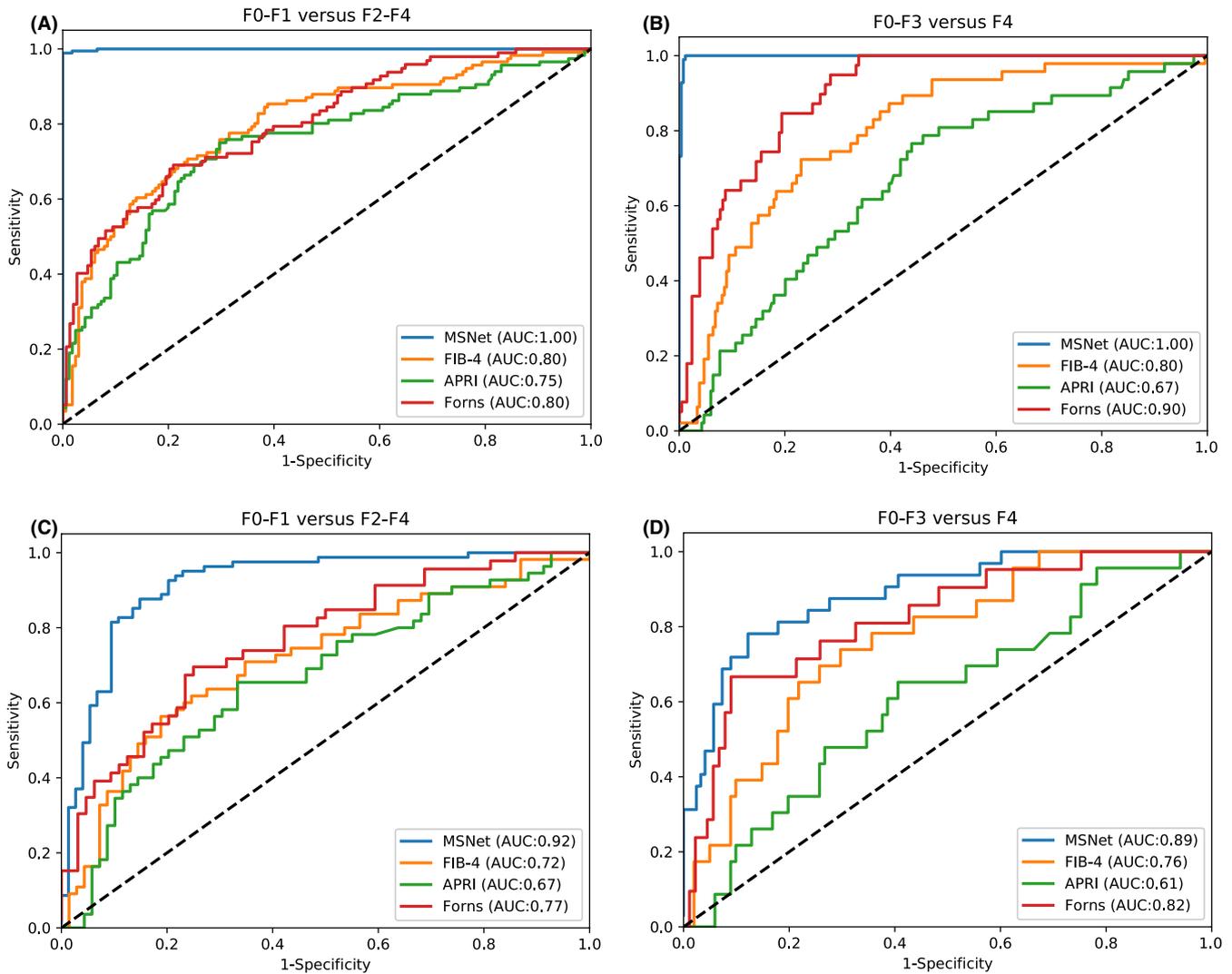


FIGURE 6 Comparison of receiver operating characteristic (ROC) curves between MSTNet, FIB-4, APRI and Forns in training and validation groups, respectively. (A, C) denotes F0-F1 versus F2-F4 ($\geq F2$) in training and validation groups. (B, D) denotes F0-F3 versus F4 (F4) in training and validation groups. AUC, area under the ROC curve; APRI, aspartate transaminase-to-platelet ratio index; FIB-4, fibrosis index based on four factors; MSTNet, multi-scale texture network. Best viewed in colour

non-fatty liver cohort (0.96 vs 0.93). Table 6 showed the corresponding diagnostic parameters of MSTNet. The sensitivity of MSTNet in fatty liver cohort reached 97.6%, which was 5.1% higher than that in non-fatty liver group. On the other hand, the specificity of MSTNet in non-fatty liver group was higher than that in fatty liver group (90.6% vs 86.5%).

3.7 | Comparison of diagnostic performance of MSTNet between sonographer crop and non-sonographer crop

We further investigated whether the diagnostic performance of MSTNet was associated with the experiences of sonographers. Three non-sonographers cropped image patches from the liver parenchyma of the ultrasound images and used this dataset as the input of MSTNet during the training and validation phase. For a

fair comparison, we ensured that two datasets had the same number of image patches and patients in the training and validation groups.

The ROC curves and diagnostic statistics were shown in Figure 9 and Table 7, respectively. We observed that the AUCs of non-sonographer crop were similar to those of sonographer crop. The AUCs were about 0.915 and 0.888 for significant fibrosis and cirrhosis, respectively. The results indicated that the diagnostic performance of MSTNet was not associated with the experiences of sonographers and MSTNet had good reproducibility.

4 | DISCUSSION

In this study, we developed a novel non-invasive approach MSTNet for the grading of liver fibrosis based on ultrasound

TABLE 3 Diagnostic performance of MSTNet, FIB-4, APRI and Forns for the assessment of liver fibrosis stages in multicentre training and validation groups

Method	G	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-
Significant fibrosis (\geq F2)								
FIB-4	T	0.80**** (0.74-0.85)	78.1 (70.9-84.1)	68.9 (59.6-77.0)	78.2 (70.9-84.0)	68.9 (59.6-77.0)	3.16 (2.73-3.66)	0.40 (0.34-0.46)
	V	0.71**** (0.62-0.81)	81.1 (69.6-89.2)	56.3 (42.4-69.4)	70.0 (58.6-72.8)	70.4 (54.6-82.7)	2.99 (2.31-3.88)	0.54 (0.42-0.70)
APRI	T	0.75**** (0.69-0.81)	70.3 (62.6-77.0)	75.0 (65.9-82.3)	80.0 (72.3-86.0)	63.9 (55.2-71.9)	2.52 (2.18-2.92)	0.35 (0.31-0.41)
	V	0.67**** (0.58-0.75)	66.6 (54.2-77.3)	65.4 (51.3-77.4)	70.7 (58.0-81.1)	61.0 (47.4-73.2)	1.96 (1.52-2.53)	0.52 (0.40-0.69)
Forns	T	0.80**** (0.74-0.86)	79.0 (71.4-85.1)	69.0 (58.7-77.8)	79.5 (71.9-85.1)	68.3 (58.1-77.1)	3.30 (2.82-3.86)	0.39 (0.33-0.46)
	V	0.77**** (0.68-0.85)	75.0 (62.3-84.6)	69.5 (54.1-81.8)	77.4 (64.7-86.7)	66.7 (51.5-79.2)	2.78 (2.19-3.53)	0.41 (0.32-0.51)
MSTNet	T	1.00 (0.99-1.00)	100.0 (97.2-100.0)	98.9 (95.8-99.8)	98.8 (95.4-98.8)	100.0 (97.4-100.0)	—	0.01 (0.01-0.01)
	V	0.92 (0.87-0.96)	85.1 (74.5-92.0)	87.6 (78.0-93.6)	86.3 (75.8-92.9)	86.6 (76.8-92.8)	5.90 (5.20-6.69)	0.14 (0.13-0.16)
Cirrhosis (F4)								
FIB-4	T	0.80**** (0.73-0.86)	76.9 (70.9-82.0)	72.3 (57.1-83.9)	93.2 (88.5-96.2)	38.6 (28.6-49.6)	3.14 (2.59-3.79)	0.36 (0.30-0.43)
	V	0.76**** (0.66-0.86)	70.2 (60.3-78.8)	73.9 (51.3-88.9)	92.2 (83.2-96.8)	36.1 (23.0-51.5)	2.49 (1.89-3.27)	0.37 (0.28-0.49)
APRI	T	0.67**** (0.57-0.77)	55.9 (49.3-62.4)	76.5 (61.6-87.2)	92.2 (86.2-95.9)	25.8 (19.0-34.1)	1.74 (1.43-2.11)	0.42 (0.34-0.51)
	V	0.61**** (0.48-0.74)	59.4 (49.2-68.9)	65.2 (42.8-82.8)	88.2 (77.6-94.4)	26.7 (16.2-40.5)	1.61 (1.14-2.26)	0.59 (0.42-0.82)
Forns	T	0.89**** (0.86-0.94)	71.3 (64.4-77.3)	94.8 (81.4-99.1)	98.6 (94.7-99.7)	38.5 (28.9-49.0)	3.31 (2.96-3.71)	0.07 (0.06-0.08)
	V	0.82*** (0.72-0.92)	91.0 (82.6-95.7)	66.6 (43.1-84.5)	77.4 (64.7-86.7)	63.6 (40.8-82.0)	7.41 (5.44-10.11)	0.37 (0.27-0.50)
MSTNet	T	1.00 (1.00-1.00)	100.0 (98.3-100.0)	100.0 (93.7-100.0)	100.0 (99.3-100.0)	100.0 (93.7-100.0)	—	0.00 (0.00-0.00)
	V	0.89 (0.83-0.95)	87.8 (80.4-92.8)	78.1 (59.6-90.1)	93.9 (87.4-97.3)	62.5 (42.8-76.8)	6.41 (5.27-7.78)	0.25 (0.20-0.30)

Note: Statistical quantifications were reported with 95% confidence interval (CI). P values were calculated by comparing AUC of MSTNet with FIB-4, APRI and Forns, respectively (*** $P < .001$; **** $P < .0001$).

Abbreviations: APRI, aspartate transaminase-to-platelet ratio index; AUC, area under the receiver operating characteristic curve; FIB-4, fibrosis index based on four factors; G, group; LR-, negative diagnostic likelihood ratio; LR+, positive diagnostic likelihood ratio; MSTNet, multi-scale texture network; NPV, negative predictive value; PPV, positive predictive value; T, training group; V, validation group.

images. We used the gold standard, liver histology to evaluate the accuracy of MSTNet and found that MSTNet was superior to FIB-4, APRI, Forns and sonographers, especially for significant fibrosis. MSTNet is easy, universally applicable and may be adopted to clinical practice.

The proposed MSTNet is a deep learning-based non-invasive liver fibrosis diagnostic approach completely based on B-mode ultrasound images. MSTNet is a liver-specific test. Furthermore, MSTNet can be embedded in B-mode ultrasound image software that can crop the image patches from the liver parenchyma and

instantly output the diagnosis results after ultrasound operation. It is very convenient to use and easy to perform. There are two studies that applied deep learning method to diagnose liver fibrosis based on B-mode ultrasound images. Meng et al³² used VGGNet to extract texture features and FCNet to achieve liver fibrosis staging. However, Meng and colleagues did not use liver biopsy, the golden standard, to root their study. Lee et al³³ developed a deep convolutional neural network (DCNN) that used a single-scale convolution kernel to predict the METAVIR score. AUCs reached 0.866 and 0.857 for \geq F2 and F4. Unlike these studies, our approach

TABLE 4 Diagnostic performance of MSTNet using different image patch resolutions on the training and validation group

Resolution	G	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-
Significant fibrosis (≥F2)								
60 × 60	T	0.98 (0.97-0.98)	91.4 (89.8-92.8)	93.0 (91.4-94.3)	93.8 (92.3-95.0)	90.4 (88.6-92.0)	10.88 (10.64-11.23)	0.07 (0.07-0.08)
	V	0.84 (0.78-0.90)	70.2 (58.3-80.1)	85.1 (75.2-91.8)	81.2 (69.2-89.5)	75.8 (65.5-83.9)	2.86 (2.41-3.41)	0.21 (0.18-0.25)
100 × 100	T	0.99 (0.98-0.99)	95.3 (90.1-97.8)	98.3 (94.9-99.6)	98.2 (94.3-99.5)	95.7 (91.5-98.0)	20.78 (19.80-21.59)	0.02 (0.02-0.02)
	V	0.90 (0.86-0.95)	74.3 (62.6-83.5)	93.8 (85.6-97.7)	91.6 (80.9-96.9)	80.0 (70.3-87.2)	3.65 (3.16-4.23)	0.08 (0.07-0.09)
120 × 120	T	0.99 (0.99-1.00)	97.0 (92.9-98.9)	97.2 (93.4-99.0)	97.0 (92.9-98.9)	97.2 (93.4-99.0)	32.88 (31.73-34.08)	0.03 (0.02-0.03)
	V	0.93 (0.89-0.97)	82.4 (71.5-89.9)	91.3 (82.5-96.2)	89.7 (79.3-95.4)	85.0 (75.4-91.5)	5.20 (4.59-5.89)	0.11 (0.09-0.12)
150 × 150	T	1.00 (0.99-1.00)	100.0 (97.2-100.0)	98.9 (95.8-99.8)	98.8 (95.4-98.8)	100.0 (97.4-100.0)	—	0.01(0.01-0.01)
	V	0.92 (0.87-0.96)	85.1 (74.5-92.0)	87.6 (78.0-93.6)	86.3 (75.8-92.9)	86.6 (76.8-92.8)	5.90 (5.20-6.69)	0.14 (0.13-0.16)
Cirrhosis (F4)								
60 × 60	T	0.97 (0.97-0.98)	94.4 (93.2-95.4)	89.1 (86.6-91.2)	95.5 (94.5-96.4)	86.6 (83.9-88.9)	16.01 (15.58-16.49)	0.11 (0.11-0.12)
	V	0.84 (0.76-0.92)	84.5 (76.7-90.2)	78.1 (59.6-90.1)	93.6 (87.0-97.2)	56.8 (41.1-71.3)	5.06 (4.15-6.16)	0.26 (0.21-0.32)
100 × 100	T	0.99 (0.99-1.00)	96.7 (93.8-98.4)	97.2 (89.4-99.5)	99.2 (97.1-99.8)	88.6 (79.0-94.3)	30.35 (28.03-31.73)	0.03 (0.03-0.03)
	V	0.88 (0.81-0.94)	83.7 (75.7-89.6)	81.2 (63.0-92.1)	94.5 (87.9-97.7)	56.5 (41.2-70.8)	4.99 (4.16-6.01)	0.22 (0.19-0.27)
120 × 120	T	0.99 (0.99-1.00)	98.5 (96.1-99.5)	100.0 (93.7-100.0)	100.0 (98.3-100.0)	94.7 (86.4-98.3)	70.25 (69.27-71.24)	0.00 (0.00-0.00)
	V	0.86 (0.78-0.94)	86.9 (79.4-92.2)	75.0 (56.2-87.8)	93.0 (86.3-96.7)	60.0 (43.4-74.7)	5.77 (4.67-7.12)	0.28 (0.23-0.36)
150 × 150	T	1.00 (1.00-1.00)	100.0 (98.3-100.0)	100.0 (93.7-100.0)	100.0 (99.3-100.0)	100.0 (93.7-100.0)	—	0.00 (0.00-0.00)
	V	0.89 (0.83-0.95)	87.8 (80.4-92.8)	78.1 (59.6-90.1)	93.9 (87.4-97.3)	62.5 (42.8-76.8)	6.41 (5.27-7.78)	0.25 (0.20-0.30)

Note: Statistical quantifications were reported with 95% confidence interval (CI). The image pyramid patches are resized to the specified resolution size and then used as an input to MSTNet.

Abbreviations: AUC, area under the receiver operating characteristic curve; G, group; LR-, negative diagnostic likelihood ratio; LR+, positive diagnostic likelihood ratio; MSTNet, multi-scale texture network; NPV, negative predictive value; PPV, positive predictive value; T, training group; V, validation group.

automatically extracts coarse-to-fine texture features through image pyramid, multi-scale convolution kernel and multi-branch attention mechanism, which results in a better diagnostic accuracy. Another advantage of MSTNet is that it is based entirely on B-mode ultrasound images and does not require additional techniques, such as SWE. Therefore, MSTNet is very convenient, inexpensive and feasible to clinical practice.

Liver biopsy is still the 'gold standard' for the assessment and quantification of liver fibrosis. However, its invasiveness causes pain and potential complications that preclude the universal applications; the sampling error and sample quality from needle biopsy specimens impact its reliability.³⁴ Scientists and clinicians have tried many ways

to explore the non-invasive approaches to substitute for liver biopsy. More than 30 non-invasive methods have been evaluated for the accuracies in comparison with liver biopsy.³⁴ However, none of these methods are satisfactory. LSM, which was designed for the evaluation of liver fibrosis, also has several shortages in clinical application. Many factors, including obesity, inflammation, cholestasis, congestion and operator experience, impact the LSM readings, even meal may affect the value of LSM. These factors might elevate the patients with F0-2 liver fibrosis to the 'cirrhotic' range of LSM and, therefore, cause misdiagnosis.³⁵ LSM is especially unreliable when the interquartile range/median ratio (IQR/M) > 30% in conjunction with LSM ≥ 7.1 kPa.³⁶ Unfortunately, 7.1 kPa is a cutoff value

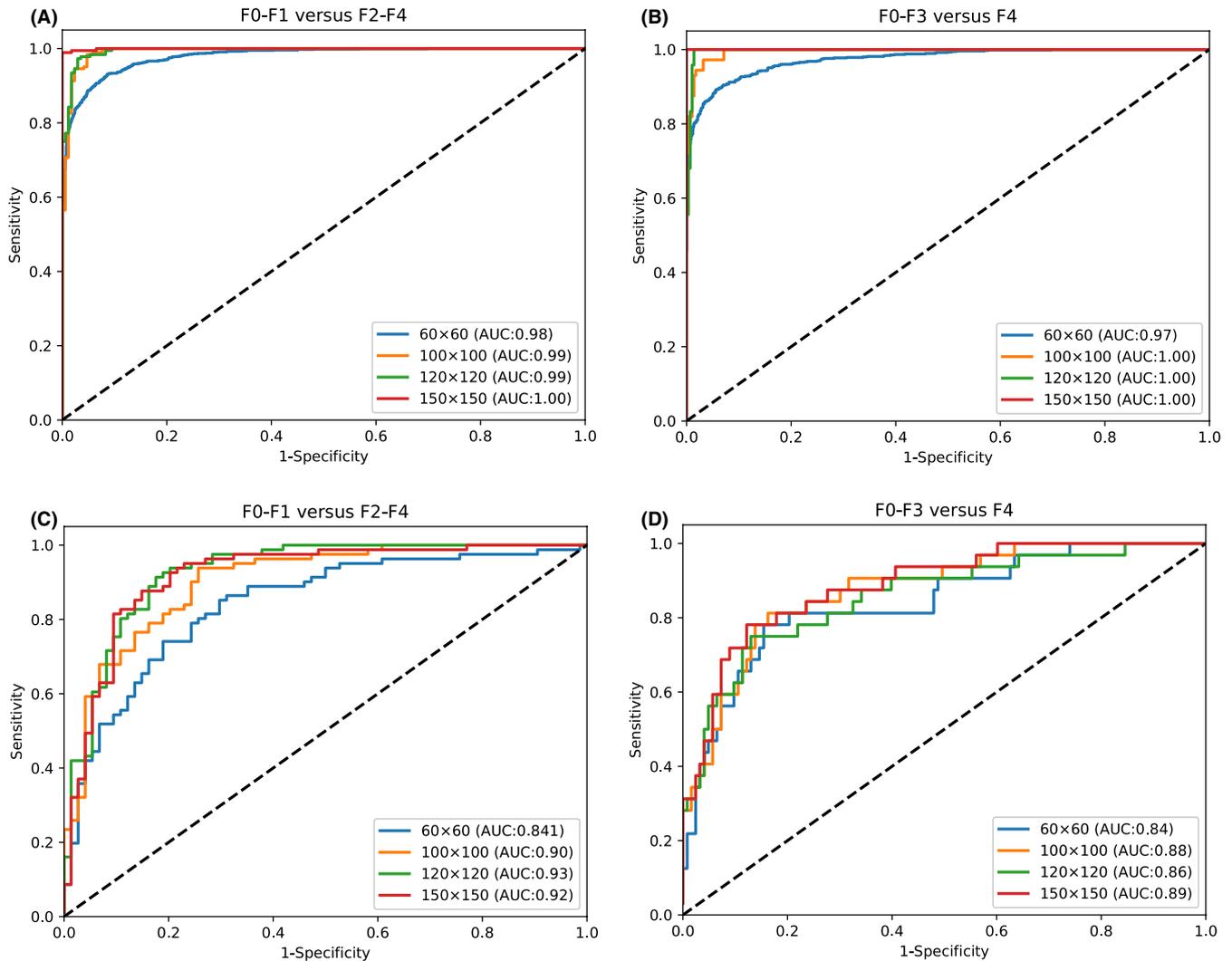


FIGURE 7 Comparison of receiver operating characteristic (ROC) curves using different image patch resolutions. (A, C) denotes F0-F1 versus F2-F4 ($\geq F2$) in training and validation groups. (B, D) denotes F0-F3 versus F4 (F4) in training and validation groups. AUC, area under the ROC curve. Best viewed in colour

TABLE 5 Diagnostic accuracy of MSTNet in comparison with three sonographers on the validation group

Sonographers	Sensitivity	Specificity	PPV	NPV	LR+	LR-
Significant fibrosis ($\geq F2$)						
Sonographer 1	80.6 (68.2-89.2)	65.2 (91.4-94.3)	75.7 (63.4-85.1)	71.4 (55.2-83.8)	3.37 (2.64-4.30)	0.43 (0.34-0.55)
Sonographer 2	85.4 (73.7-92.7)	58.7 (43.3-72.7)	73.6 (61.7-83.0)	75.0 (57.5-87.3)	4.04 (3.11-5.26)	0.48 (0.37-0.63)
Sonographer 3	82.2 (70.0-90.4)	78.3 (63.2-88.5)	83.6 (71.4-91.4)	76.6 (61.6-87.2)	4.41 (3.64-5.34)	0.26 (0.22-0.32)
MSTNet	85.1 (74.5-92.0)	87.6 (78.0-93.6)	86.3 (75.8-92.9)	86.6 (76.8-92.8)	5.90 (5.20-6.69)	0.14 (0.13-0.16)

Note: Statistical quantifications were reported with 95% confidence interval (CI).

Abbreviations: LR-, negative diagnostic likelihood ratio; LR+, positive diagnostic likelihood ratio; MSTNet, multi-scale texture network; NPV, negative predictive value; PPV, positive predictive value.

between F1 and F2 in clinical practice, and therefore, the LSM value may not be used to guide the treatments. Furthermore, the grading of fibrosis according to LSM is not universally standardized. Castera

et al³⁷ defined 7.1 kPa for $F \geq 2$, 9.5 kPa for $F \geq 3$ and 12.5 kPa for $F = 4$, while Ziolk et al³⁸ defined ≥ 8.8 kPa for $F \geq 2$ and 14.6 kPa for $F = 4$.

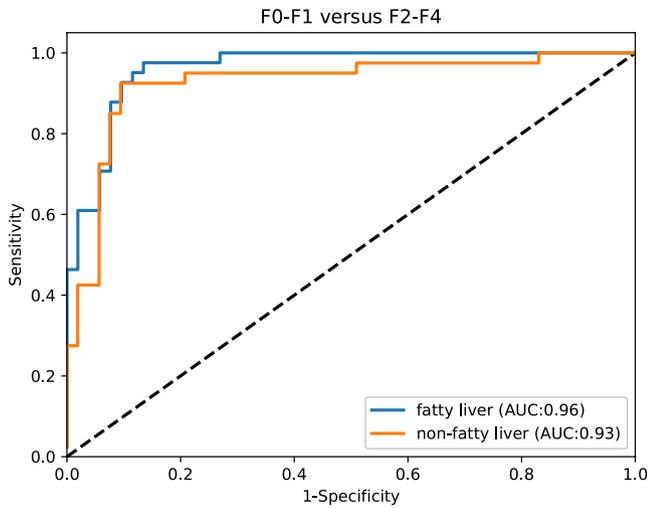


FIGURE 8 Comparison of receiver operating characteristic (ROC) curves of MSTNet in patients with and without fatty liver for the assessment of significant fibrosis ($\geq F2$). AUC, area under the ROC curve; MSTNet, multi-scale texture network. Best viewed in colour

Our approach seemed to solve the problems. MSTNet capture multi-scale texture features of ultrasound images that conquered the sampling errors in liver biopsy. MSTNet was not affected by liver inflammation and fat content. Furthermore, the diagnostic accuracy was not influenced by the experience of sonographer. The resolution of the input image patch did not impact the accuracy as long as it is bigger than 100×100 . Moreover, because of the involvement of the computer, the value of MSTNet is more objective.

Our study has some limitations. Firstly, the unbalanced patient data distribution leaves much room for improvement in the diagnostic accuracy of cirrhosis. In our data distribution, the number of patients with normal and mild fibrosis (F0-F1) is about twice of the patients with cirrhosis (F4), which makes the model tend to learn the texture features of F0-F1. As a result, MSTNet demonstrates relatively lower diagnostic accuracy for F4. Second, the data volume of multi-centre patients needs to be expended. Third, although the experience of sonographer does not affect the diagnostic accuracy, the effect of different ultrasound instruments on the diagnostic accuracy needs further study. Fourth, we did

TABLE 6 Diagnostic performance of MSTNet in patients with and without fatty liver

Fatty liver	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-
Significant fibrosis ($\geq F2$)							
	0.96 (0.93-0.99)	97.6 (85.6-99.9)	86.5 (73.6-94.0)	85.1 (71.1-93.3)	97.8 (87.0-99.9)	7.24 (3.63-14.46)	0.03 (0.01-0.20)
	0.93 (0.86-0.98)	92.5 (78.5-98.0)	90.6 (78.6-96.5)	88.1 (73.6-95.5)	94.1 (82.8-98.5)	9.81 (4.24-22.68)	0.08 (0.02-0.25)

Note: Statistical quantifications were reported with 95% confidence interval (CI).

Abbreviations: LR-, negative diagnostic likelihood ratio; LR+, positive diagnostic likelihood ratio; MSTNet, multi-scale texture network; NPV, negative predictive value; PPV, positive predictive value.

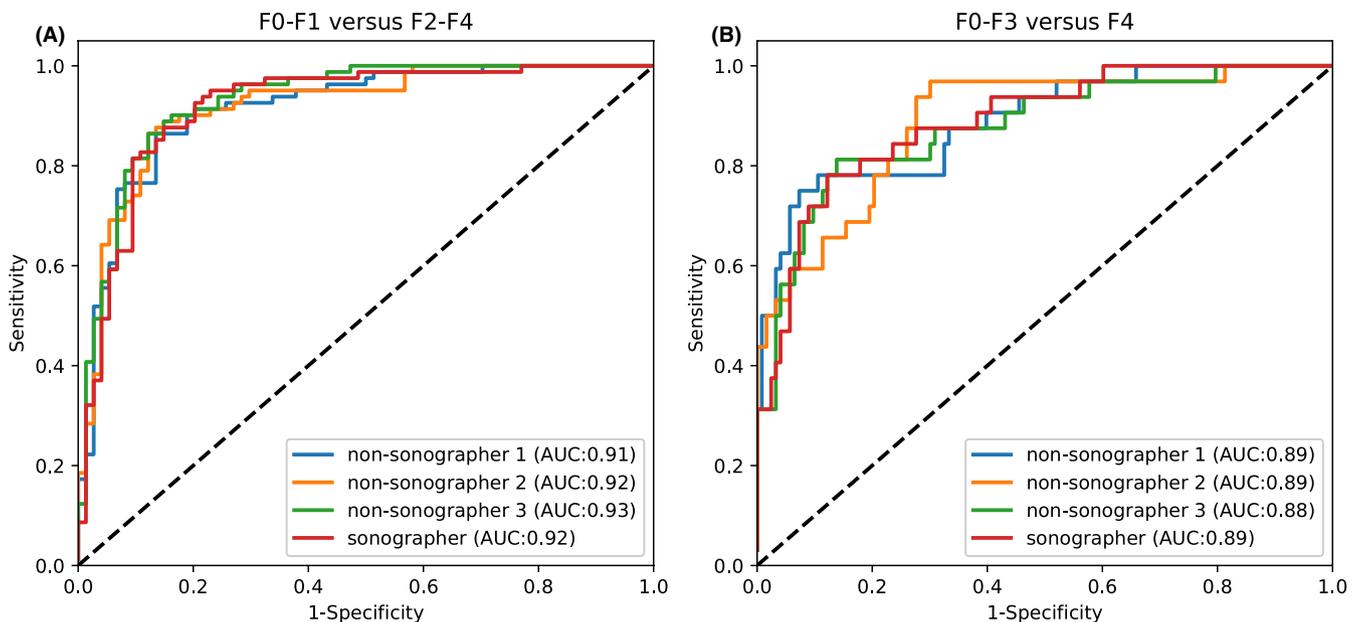


FIGURE 9 Comparison of receiver operating characteristic (ROC) curves of MSTNet between sonographer crop and non-sonographer crop. (A) denotes F0-F1 versus F2-F4 ($\geq F2$) in validation group. (B) denotes F0-F3 versus F4 (F4) in validation group. AUC, area under the ROC curve; MSTNet, multi-scale texture network. Best viewed in colour

TABLE 7 Comparison of diagnostic performance of MSTNet between sonographer crop and non-sonographer crop on the validation group

Crop	AUC	Sensitivity	Specificity	PPV	NPV	LR+	LR-
Significant fibrosis (\geq F2)							
Son	0.92 (0.87-0.96)	85.1 (74.5-92.0)	87.6 (78.0-93.6)	86.3 (75.8-92.9)	86.6 (76.8-92.8)	5.90 (5.20-6.69)	0.14 (0.13-0.16)
NSon 1	0.91 (0.87-0.96)	86.4 (76.5-92.7)	86.5 (76.1-92.9)	87.5 (77.8-93.5)	85.3 (74.8-92.1)	6.39 (3.57-11.45)	0.16 (0.09-0.27)
NSon 2	0.92 (0.87-0.96)	86.4 (76.5-92.7)	87.8 (77.6-93.9)	88.6 (79.0-94.3)	85.5 (75.1-92.2)	7.11 (3.83-13.19)	0.15 (0.09-0.27)
NSon 3	0.93 (0.89-0.97)	86.4 (76.5-92.7)	87.8 (77.6-93.9)	88.6 (79.0-94.3)	85.5 (75.1-92.2)	7.11 (3.83-13.19)	0.15 (0.09-0.27)
Cirrhosis (F4)							
Son	0.89 (0.83-0.95)	87.8 (80.4-92.8)	78.1 (59.6-90.1)	93.9 (87.4-97.3)	62.5 (42.8-76.8)	6.41 (5.27-7.78)	0.25 (0.20-0.30)
NSon 1	0.89 (0.82-0.96)	75.0 (56.2-87.8)	92.7 (86.2-96.4)	72.7 (54.2-86.1)	93.4 (87.1-96.9)	10.25 (5.29-19.83)	0.27 (0.14-0.49)
NSon 2	0.89 (0.82-0.95)	96.8 (82.0-99.8)	69.9 (60.9-77.7)	45.6 (33.6-58.0)	98.9 (92.8-99.9)	3.22 (2.44-4.25)	0.04 (0.01-0.31)
NSon 3	0.88 (0.81-0.95)	81.3 (62.9-92.1)	86.2 (78.5-91.5)	60.5 (44.5-74.6)	94.6 (88.2-97.8)	5.87 (3.67-9.42)	0.22 (0.11-0.45)

Note: Statistical quantifications were reported with 95% confidence interval (CI).

Abbreviations: LR-, negative diagnostic likelihood ratio; LR+, positive diagnostic likelihood ratio; MSTNet, multi-scale texture network; NPV, negative predictive value; NSon, non-sonographer crop; PPV, positive predictive value; Son, sonographer crop.

not compare MSTNet with LSM, class I (direct) biomarkers of liver fibrosis, such as M2BPGi and ELF score. However, these non-invasive approaches also used liver histology as gold standard. The accuracy of MSTNet was rooted on liver histology, and therefore, our study delivered a clear message. In the end, we only use the liver parenchyma texture of ultrasound images to diagnose liver fibrosis. To some extent, liver capsule morphology change may also help assess liver fibrosis, especially for cirrhosis, which will be explored in our future study.

In conclusion, our study demonstrated that MSTNet outperformed three serum biomarkers in assessing significant fibrosis and cirrhosis. In particular, the diagnostic accuracy of MSTNet was not impacted by fat content in the liver. Furthermore, MSTNet was more accurate than sonographers in the diagnosis of significant fibrosis. These results demonstrated that MSTNet had an excellent potential for clinical application.

ACKNOWLEDGEMENTS

The work was supported by the 13-5 State S&T Projects of China (2018ZX10302-206) and the Natural Science Foundation of China (81871646 and 81870425). We thank Hongqun Liu for the assistance during the revision of this manuscript. Dongsheng Ruan, Yu Shi and Linfeng Jin contributed equally to this work.

CONFLICT OF INTEREST

The authors disclose no conflict of interest.

ORCID

Dongsheng Ruan  <https://orcid.org/0000-0001-6286-2245>

Yu Shi  <https://orcid.org/0000-0002-9392-7176>

Weiyang Zheng  <https://orcid.org/0000-0002-8011-0503>

Min Zheng  <https://orcid.org/0000-0001-6159-9879>

REFERENCES

- Bataller R, Brenner DA. Liver fibrosis. *J Clin Investig.* 2005;115(2):209-218.
- Bravo AA, Sheth SG, Chopra S. Liver biopsy. *N Engl J Med.* 2001;344(7):495-500.
- Guéchet J, Loria A, Serfaty L, Giral P, Giboudeau J, Poupon R. Serum hyaluronan as a marker of liver fibrosis in chronic viral hepatitis C: effect of α -interferon therapy. *J Hepatol.* 1995;22:1.
- Shi Y, Zheng M. Hepatitis B virus persistence and reactivation. *BMJ.* 2020;370.
- Bateson MC, Bouchier IAD. *Clinical Investigations in Gastroenterology.* Kluwer Academic Publishers; 1988.
- Rockey DC, Caldwell SH, Goodman ZD, et al. Liver biopsy. *Hepatology.* 2009;49(3):1017-1044.
- Guéchet J, Laudat A, Loria A, et al. Diagnostic accuracy of hyaluronan and type III procollagen amino-terminal peptide serum assays as markers of liver fibrosis in chronic viral hepatitis C evaluated by ROC curve analysis. *Clin Chem.* 1996;42(4):558-563.
- Griffiths C, Rooney C, Brock A. Leading causes of death in England and Wales—how should we group causes. *Health Stat Q.* 2005;28(9):6-17.
- Bosetti C, Levi F, Lucchini F, et al. Worldwide mortality from cirrhosis: an update to 2002. *J Hepatol.* 2007;46(5):827-839.
- Oberti F, Valsesia E, Pilette C, et al. Noninvasive diagnosis of hepatic fibrosis or cirrhosis. *Gastroenterology.* 1997;113(5):1609-1616.
- Afdhal NH, Nunes D. Evaluation of liver fibrosis: a concise review. *Am J Gastroenterol.* 2004;99(6):1160-1174.
- Olaso E, Friedman SL. Molecular regulation of hepatic fibrogenesis. *J Hepatol.* 1998;29(5):836-847.
- Bonacini M, Hadi G, Govindarajan S, et al. Utility of a discriminant score for diagnosing advanced fibrosis or cirrhosis in patients with chronic hepatitis C virus infection. *Am J Gastroenterol.* 1997;92(8):1302-1304.
- Wang K, Lu X, Zhou H, 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(4):729-741.
- Afdhal NH. Biopsy or biomarkers: is there a gold standard for diagnosis of liver fibrosis? *Clin Chem.* 2004;50(8):1299-1300.
- Fabris C, Federico E, Soardo G, et al. Blood lipids of patients with chronic hepatitis: differences related to viral etiology. *Clin Chim Acta.* 1997;261(2):159-165.
- Loeza-del-Castillo A, Paz-Pineda F, Oviedo-Cárdenas E, et al. AST to platelet ratio index (APRI) for the noninvasive evaluation of liver fibrosis. *Ann Hepatol.* 2008;7(4):350-357.
- Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection comparison with liver biopsy and fibrotest. *Hepatology.* 2007;46(1):32-36.
- Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. *J Hepatol.* 2007;47(4):598-607.
- Ganne-Carrié N, Ziol M, de Ledinghen V, et al. Accuracy of liver stiffness measurement for the diagnosis of cirrhosis in patients with chronic liver diseases. *Hepatology.* 2006;44(6):1511-1517.
- Mueller S, Sandrin L. Liver stiffness: a novel parameter for the diagnosis of liver disease. *Hepat Med.* 2010;2:49.
- Ferraioli G, Tinelli C, Dal Bello B, et al. Performance of liver stiffness measurements by transient elastography in chronic hepatitis. *World J Gastroenterol.* 2013;19(1):49.
- Wong G-H, Wong V-S, Chim A-L, et al. Factors associated with unreliable liver stiffness measurement and its failure with transient elastography in the Chinese population. *J Gastroenterol Hepatol.* 2011;26(2):300-305.
- Ghany MG, Doo E. Assessment of liver fibrosis: palpate, poke or pulse? *Hepatology.* 2005;42(4):759-761.
- Ferraioli G, Tinelli C, Dal Bello B, et al. Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. *Hepatology.* 2012;56(6):2125-2133.
- Ogawa K, Fukushima M, Kubota K, et al. Computer-aided diagnostic system for diffuse liver diseases with ultrasonography by neural networks. *IEEE Trans Nucl Sci.* 1998;45(6):3069-3074.
- Wu CM, Chen YC, Hsieh KS. Texture features for classification of ultrasonic liver images. *IEEE Trans Med Imaging.* 1992;11(2):141-152.
- Yeh WC, Huang SW, Li PC. Liver fibrosis grade classification with B-mode ultrasound. *Ultrasound Med Biol.* 2003;29(9):1229-1235.
- Mojsilovic A, Popovic M, Markovic S, et al. Characterization of visually similar diffuse diseases from B-scan liver images using nonseparable wavelet transform. *IEEE Trans Med Imaging.* 1998;17(4):541-549.
- Li W, Huang Y, Zhuang B-W, et al. Multiparametric ultrasonics of significant liver fibrosis: a machine learning-based analysis. *Eur Radiol.* 2019;29(3):1496-1506.
- Wu K, Chen X, Ding M. Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. *Optik.* 2014;125(15):4057-4063.

32. Meng D, Zhang L, Cao G, et al. Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. *IEEE Access*. 2017;5:5804-5810.
33. Lee JH, Joo I, Kang TW, et al. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. *Eur Radiol*. 2020;30(2):1264-1273.
34. Chin JL, Pavlides M, Moolla A, et al. Non-invasive markers of liver fibrosis: Adjuncts or alternatives to liver biopsy? *Frontiers in Pharmacology*. 2016;7:159
35. Patel K, Sebastiani G. Limitations of non-invasive tests for assessment of liver fibrosis. *JHEP Reports*. 2020;2(2):100067.
36. Boursier J, Zarski J-P, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology*. 2013;57(3):1182-1191.
37. Castéra L, Vergniol J, Foucher J, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology*. 2005;128(2):343-350.
38. Ziol M, Handra-Luca A, Kettaneh A, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. *Hepatology*. 2005;41(1):48-54.

How to cite this article: Ruan D, Shi Y, Jin L, et al. An ultrasound image-based deep multi-scale texture network for liver fibrosis grading in patients with chronic HBV infection. *Liver Int*. 2021;41:2440–2454. <https://doi.org/10.1111/liv.14999>