RESEARCH

Liver fibrosis stage classification in stacked microvascular images based on deep learning

Daisuke Miura[1,2], Hiromi Suenaga[2*], Rino Hiwatashi[1] and Shingo Mabu[3]

Abstract

Background   Monitoring fibrosis in patients with chronic liver disease (CLD) is an important management strategy. We have already reported a novel stacked microvascular imaging (SMVI) technique and an examiner scoring evaluation method to improve fibrosis assessment accuracy and demonstrate its high sensitivity. In the present study, we analyzed the effectiveness and objectivity of SMVI in diagnosing the liver fibrosis stage based on artificial intelligence (AI).

Methods   This single-center, cross-sectional study included 517 patients with CLD who underwent ultrasonography and liver stiffness testing between August 2019 and October 2022. A convolutional neural network model was constructed to evaluate the degree of liver fibrosis from stacked microvascular images generated by accumulating high-sensitivity Doppler (i.e., high-definition color) images from these patients. In contrast, as a method of judgment by the human eye, we focused on three hallmarks of intrahepatic microvessel morphological changes in the stacked microvascular images: narrowing, caliber irregularity, and tortuosity. The degree of liver fibrosis was classified into five stages according to etiology based on liver stiffness measurement: F0–1Low (< 5.0 kPa), F0–1High (≥ 5.0 kPa), F2, F3, and F4.

Results   The AI classification accuracy was 53.8% for a 5-class classification, 66.3% for a 3-class classification (F0–1Low vs. F0–1High vs. F2–4), and 83.8% for a 2-class classification (F0–1 vs. F2–4). The diagnostic accuracy for ≥ F2 was 81.6% in the examiner's score assessment, compared with 83.8% in AI assessment, indicating that AI achieved higher diagnostic accuracy. Similarly, AI demonstrated higher sensitivity and specificity of 84.2% and 83.5%, respectively. Comparing human judgement with AI judgement, the AI analysis was a superior model with a higher F1 score in the 2-class classification.

Conclusions   In detecting significant fibrosis (≥ F2) using the SMVI method, AI-based assessments are more accurate than human judgement; moreover, AI-based SMVI analysis eliminating human subjectivity bias and determining patients with objective fibrosis development is considered an important improvement.

Keywords   Artificial intelligence, Deep learning, Liver cirrhosis, Microvascular imaging, Stacked microvascular imaging

*Correspondence:
Hiromi Suenaga
hiro1204@yamaguchi-u.ac.jp
Full list of author information is available at the end of the article

Miura *et al. BMC Medical Imaging* (2025) 25:8

Page 2 of 11

Introduction

Chronic liver disease (CLDs) is a chronic inflammation of the liver caused by various factors, resulting in persistent liver dysfunction [1]. In recent years, the increase in chronic hepatitis caused by lifestyle habits, such as excess alcohol consumption and metabolic syndrome, has become a worldwide problem [2]. Liver fibrosis associated with chronic inflammation is considered a critical factor in predicting the development of liver-related complications, such as hepatic failure and carcinogenesis [3].

Although liver biopsy is crucial in assessing liver fibrosis, its invasiveness, low reproducibility, sampling bias, and inconsistent pathological results are disadvantages [4]. Ultrasound elastography has recently received attention as a non-invasive diagnostic test for fibrosis that is more suitable than liver biopsy for routine practice [5, 6]. Magnetic resonance elastography is considered to be the most accurate noninvasive method of detecting advanced fibrosis but is not recommended in clinical practice due to its high cost and limited use. Transient elastography, on the other hand, is a validated technique and the most widely available device that is easy to use in clinical practice. Transient elastography and ultrasound-based measurements such as point-shear wave elastography (SWE) and 2D-SWE have an applicability of > 95% [7]. With this advancement, the European Association for the Study of the Liver (EASL) states in its latest 2021 guidelines that advanced liver fibrosis can be diagnosed without liver biopsy when liver stiffness measurement (LSM) with vibration-controlled transient elastography (VCTE), and a patented blood test [7]. However, LSM obtained by VCTE may overestimate the liver fibrosis stage due to acute inflammation, cholestasis, and congestion [8]. To overcome these challenges, a previous study developed stacked microvascular imaging (SMVI), a novel Doppler imaging technique that can depict details of intrahepatic vascular changes caused by fibrosis [9]. In addition, the study proposed a new liver fibrosis assessment method using an SMVI scoring system based on three vascular hallmarks (narrowing, caliber irregularity, and tortuosity), which demonstrated good inter- and intra-examiner reliability with weighted kappa coefficients ranging from 0.72 to 0.89 and high fibrosis staging ability [9]. On the other hand, the inter- and intra-examiner reliability for VCTE using intraclass correlation coefficients ranged from 0.79 to 0.84, while the inter- and intra-examiner reliability for 2D-SWE is reported to range from 0.71 to 0.85 [10]. Hence, the reliability of SMVI is considered to be almost equivalent to that of elastography. In particular, SMVI yielded higher diagnostic accuracy than other evaluation methods, particularly in the early stage of fibrosis. However, it has the disadvantage of being dependent on subjective human judgment. Thus, a more objective evaluation method is required.

In recent years, computer image analysis using artificial intelligence (AI), particularly deep learning, has increasingly and widely been applied in the field of medical imaging. Several studies of ultrasound image classification using convolutional neural networks (CNN) in thyroid [11, 12], breast [13], soft tissue [14], and liver tumors [15] have been published. Some studies that have attempted to classify liver fibrosis staging by deep learning image analysis and have reported accuracies of 94%, 83–88%, and 88% for ultrasound [16], computed tomography [17], and magnetic resonance imaging [18], respectively. Deep learning with CNNs is based on a multi-step process in which the computer automatically learns higher-order image features, extracts those features, and classifies them using fully connected layers in the network, without the need for a human to design the features.

Here, we trained a CNN model on 517 stacked microvascular images to validate the accuracy of AI-based liver fibrosis stage classification and compared its accuracy with that of a human scoring method.

Materials and methods

Ethics consideration

This study was reviewed and approved by the Fukuoka Tokushukai Hospital Institutional Review Board (Approval Number, 220101), and informed consent was obtained from all patients included in the study.

Study sample

Stacked microvascular images were obtained from 564 patients with suspected CLD who underwent VCTE (FibroScan) of the liver between August 2019 and October 2022 at Fukuoka Tokushukai Hospital (Fukuoka, Japan). Patients with the following conditions affecting LSM were excluded: (1) multiple or large hepatic masses, (2) acute liver injury, (3) heart failure, (4) severe fatty liver disease, and (5) inability to hold breath. FibroScan was not performed for cases with jaundice, perihepatic ascites, or pregnancy. Thirty-one healthy individuals without CLD were excluded, and 517 patients with CLD were enrolled (Fig. 1).

Reference standard

The EASL guidelines state that liver biopsy is not required for the diagnosis of liver fibrosis, with transient elastography and shear wave elastography being agreed upon as alternatives [7]. In fact, LSM-based fibrosis staging was adopted in this study because liver biopsy for fibrosis staging for CLD has been replaced with elastography by FibroScan in general hospitals.

```
┌──────────────────────────────────────────────────────────────┐
│ Inclusion Criterion                                            │
│ Patients (≥ 20 years) who underwent liver elastography (n = 564)│
└──────────────────────────────────────────────────────────────┘
        │
        │         ┌──────────────────────────────────────────────┐
        ├────────▶│ Excluded (n = 0)                             │
        │         │ Peri-hepatic ascites (n = 0)                 │
        │         │ Cholestatic icterus (n = 0)                  │
        │         │ Pregnancy (n = 0)                            │
        │         │ FibroScan® was not performed on these patients.│
        │         └──────────────────────────────────────────────┘
        ▼
┌──────────────────┐
│ 564 patients     │
└──────────────────┘
        │
        │         ┌──────────────────────────────────────────────┐
        ├────────▶│ Excluded (n = 31)                            │
        │         │ Healthy subjects (n = 31)                    │
        │         └──────────────────────────────────────────────┘
        ▼
┌──────────────────┐
│ 533 patients     │
└──────────────────┘
        │
        │         ┌───────────────────────────────────────────────────────────────┐
        ├────────▶│ Excluded (n = 16)                                             │
        │         │ Presence of multiple or large liver tumors (n = 3)            │
        │         │ Acute liver injury (ALT > 200 U/L) (n = 2)                    │
        │         │ Heart failure (n = 2)                                         │
        │         │ Severe steatosis (n = 5)                                      │
        │         │ Inability to stop breathing on request during examination (n = 4)│
        │         └───────────────────────────────────────────────────────────────┘
        ▼
┌──────────────────────────────────────────────────────────────────┐
│ Enrolled (n = 517)                                                │
│ Chronic liver disease (n = 517)                                   │
│   Chronic HCV infection (n = 167)                                 │
│   Chronic HBV infection (n = 126)                                 │
│   Metabolic dysfunction-associated fatty liver disease (n = 90)   │
│   MetALD (n = 7)                                                   │
│   Cryptogenic SLD (n = 4)                                          │
│   Alcoholic liver disease (n = 78)                                │
│   Autoimmune hepatitis (n = 14)                                   │
│   Primary biliary cholangitis (n = 12)                            │
│   Drug (n = 5)                                                     │
│   Unknown etiology (n = 14)                                       │
└──────────────────────────────────────────────────────────────────┘
```

Fig. 1 Flowchart of the participants. ALD, alcohol-associated liver disease; HBV, hepatitis B virus; HCV, hepatitis C virus; SLD, steatotic liver disease

The etiology-dependent staging of fibrosis was determined from the LSM values, as defined by ECHOSENS™ as follows: Chronic hepatitis C: F0-F1: ≤ 7.1 kPa; F2: $7.2 - 9.4$; F3: $9.5 - 12.4$; F4: ≥ 12.5; Chronic hepatitis B: F0-F1: ≤ 7.1 kPa; F2: $7.2 - 9.3$; F3: $9.4 - 12.1$; F4: ≥ 12.2; Metabolic dysfunction associated steatotic liver disease: F0-F1: ≤ 6.9 kPa; F2: $7.0 - 8.6$; F3: $8.7 - 10.2$; F4: ≥ 10.3; Alcohol-associated liver disease: F0-F1: ≤ 8.9 kPa; F2: $9.0 - 12.0$; F3: $12.1 - 18.5$; F4: ≥ 18.6; Primary biliary cholangitis: F0-F1: ≤ 8.7 kPa; F2: $8.8 - 10.6$; F3: $10.7 - 16.7$; F4: ≥ 16.8; and Autoimmune hepatitis: F0-F1: ≤ 5.8 kPa; F2: $5.9 - 10.5$; F3: $10.6 - 16.0$; F4 ≥ 16.1. Since FibroScan cannot clearly discriminate between the F0 and F1 stages, we used the upper limit of normal (5.0 kPa) as the cut-off value and designated LSM values of < 5.0 as F0–1Low and ≥ 5.0 as F0–1High [19].
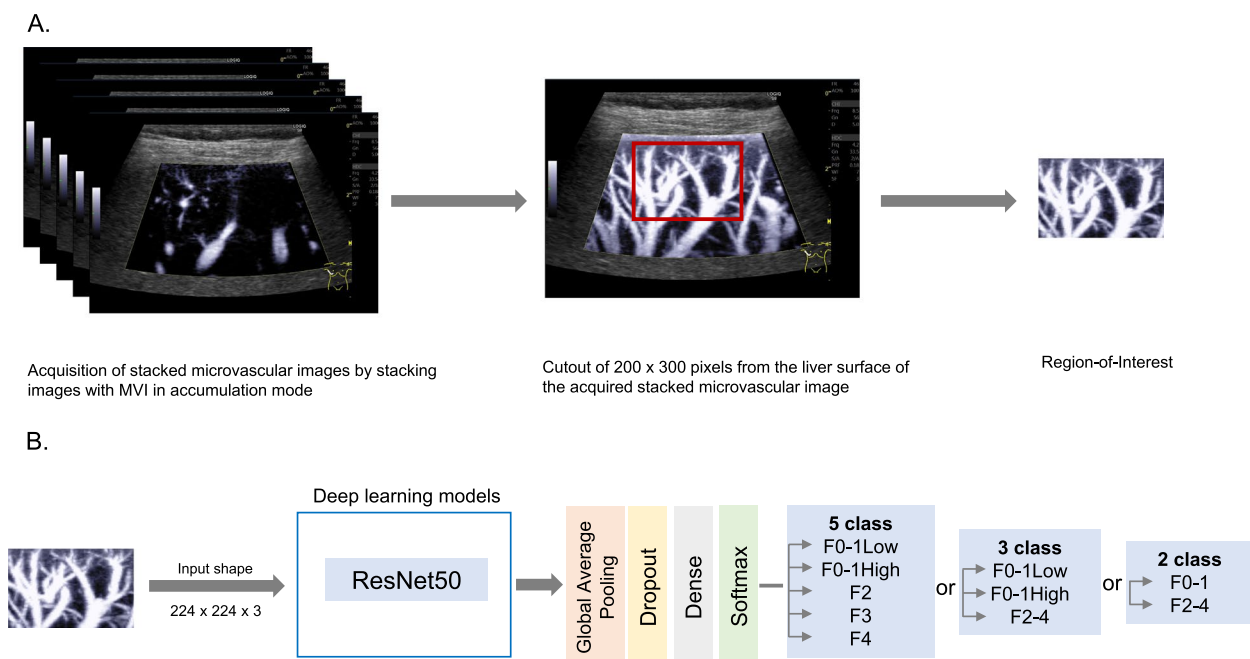
Data preprocessing and designing a convolutional neural network model

The ultrasound equipment used to obtain stacked microvascular images was a Logiq S8 FS with a 9-L probe (GE Healthcare, Chicago, IL, USA). The protocol for obtaining stacked microvascular images was based on the previously developed method [9]. Specifically, the probe

Miura *et al. BMC Medical Imaging*   (2025) 25:8

Page 4 of 11

was positioned in the right intercostal space where the liver surface could be observed; the high-definition color was set to infinite accumulation; and a tilt scan was performed once, while holding the breath for 3–5 s.

To visualize intrahepatic microvascular morphological changes in early fibrosis, stacked microvascular images were targeted to the liver surfaces. These images were cropped into rectangles (200×300 pixels) containing the liver surface nadir in the colored region of interest by one investigator. This cropping ensured that color Doppler artifacts were eliminated. In general, deep learning models exhibit high predictive performance by inputting large amounts of training data and learning the features contained in the data. However, SMVI is a new technique not yet widely implemented, and it is difficult to obtain a large dataset; therefore, data augmentation was performed. We implemented data augmentation techniques, such as rotation range, vertical flip, horizontal flip, zoom, width shift, and height shift, using the ImageDataGenerator implemented in TensorFlow developed by Google (Mountain View, CA, USA). The data preprocessing steps are illustrated in Fig. 2A. The CNN backbone model used in this study is ResNet50, introduced in 2015 by Kaming et al. [20]. The model achieves high accuracy by introducing a residual block that prevents gradient loss even when the layers are deepened. In the training, ResNet50, which has been widely used in the deep learning field, was used. This ResNet50 architecture, which is shown in Fig. 3, has the same architecture as the original ResNet paper [20], where 16 residual blocks (totally 49 layers) and one fully-connected layer are connected. The optimizer used is Adam, which combines momentum and further suppresses the oscillations by moving averages, and RMSProp, which suppresses the oscillations by adjusting the learning rate. In addition, sparse categorical cross-entropy was used as the loss function because the labels were integers. As the number of participants in the dataset was disproportionate, the optimization was weighted by the inverse of the class number ratio. Specifically, the weights for each class were calculated as follows: class_weights = total_count / class_count. By adapting these weights to the loss function, the effect of class imbalance was reduced. All the 517 images were split into 10 subsets and 10-Fold Cross-Validation was implemented for performance evaluation. In the training, the number of epochs was set to 70, which was experimentally determined by considering sufficient training of ResNet50. In the testing, the trained ResNet50 was applied to the test data for evaluating the accuracy, precision, recall, and F1 score. A conceptual diagram of these CNN designs is presented in Fig. 2B.

A.

Acquisition of stacked microvascular images by stacking images with MVI in accumulation mode

Cutout of 200 x 300 pixels from the liver surface of the acquired stacked microvascular image

Region-of-Interest

B.

Fig. 2 Stacked microvascular imaging data preprocessing and designing a convolutional neural network model. **A** Stacked microvascular image generation. Stacked microvascular images generated by accumulating the high-sensitivity Doppler mode called high-definition color for a few seconds can depict many intrahepatic vessels without losing their continuity. To avoid artifacts, 200×300 pixels were cropped from the Doppler Region-of-Interest. **B** Convolutional neural network model construction. The cropped stacked microvascular images were learned by ResNet50

layer name	output size	50-layer
conv1	122×112	7×7, 64, stride 2
conv2_x	56×56	3×3 max pool, stride 2
		$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax
FLOPs		3.8×10^9

Fig. 3 Architectures for ImageNet. ResNet50 contains 16 residual blocks (49 layers in total), with one fully connected layer

Judgment by a sonographer using the SMVI scoring system

As a method of judgment by the human eye, we focused on three hallmarks of intrahepatic microvessel morphological changes in the stacked microvascular images: narrowing, caliber irregularity, and tortuosity. Each hallmark was scored from 0 to 2 based on severity: 0 = absent, 1 = mild, and 2 = present [9].

"Narrowing" was determined based on the average diameter of the five largest vessels located 1 cm deep from the liver surface: 0 for ≥ 1.50 mm, 1 for 1.25–1.49 mm, and 2 for ≤ 1.24 mm. "Caliber irregularity" was defined as minute variations in vessel diameter and was determined by identifying the regularity/irregularity of vessel contours within 1.5 cm of the liver surface. The frequency of caliber irregularity was scored as follows: 0, no irregularity in any vessel; 1, irregularity in < 50% of the vessels; and 2, irregularity in ≥ 50% of the vessels. "Tortuosity" is defined by the straightness/winding of the depicted vessel. The vessels within 1.5 cm from the liver surface were carefully observed and scored: 0, no winding in any vessel; 1, winding in < 50% of the vessels; and 2, winding in ≥ 50% of the vessels. The sum of the three scores was defined as the SMVI score, ranging between 0 and 6. One sonographer acquired the stacked microvascular images, and another sonographer, who did not have access to all clinical information, independently performed the scoring judgment for all cases. The sonographer who acquired the SMVI images has 15 years' experience, and the sonographer who made the judgement has 10 years' experience; both sonographers are certified by the Japan Society of Ultrasonics in Medicine.

A comparison of representative stacked microvascular images of deferent fibrosis stages is shown in Fig. 4.

Statistical analyses

The Shapiro–Wilk test was used to test the normality of the data, with a significance level of $p < 0.05$. Validation of AI image discriminability in different fibrosis stages was conducted in three conditions: 2-classes (F0–1 and F2–4), 3-classes (F0–1Low, F0–1High, and F2–4), and 5-classes (F0–1Low, F0–1High, F2, F3, and F4). The diagnostic accuracy of the SMVI score to distinguish the significant fibrosis (≥ F2) from the other stages was evaluated as the area under the receiver operating characteristic (ROC) curve (AUC), and cutoffs were determined based on the Youden index. All statistical analyses were performed using EZR [21] (Saitama Medical Center, Jichi Medical University, Saitama, Japan).

Results

Patient population

In this retrospective, cross-sectional study, we enrolled 517 consecutive patients (mean age, 63.9 ± 12.9 years; 254 [49.1%] males, 263 [50.9%] females) with mixed CLD. The demographic information of these patients is summarized in Table 1. In 56.7% of patients with CLD, the etiology involved viral liver disease caused by hepatitis B virus and hepatitis C virus, and in 34.6%, the etiology was steatotic liver disease. Significant fibrosis (≥ F2) was
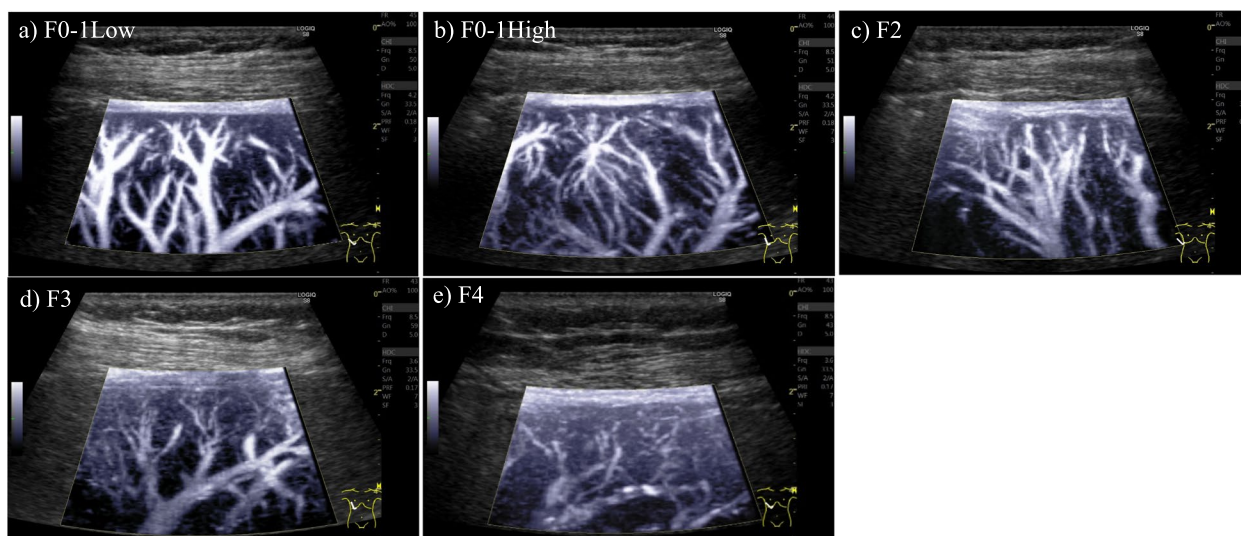
Fig. 4 Comparison of the representative stacked microvascular images in different fibrosis stages. With the development of fibrosis, complex modifications of the vessels occur such as narrowing, caliber irregularities, and tortuosity. The scores for each fibrosis stage are as follows in order of narrowing, irregular caliber, and tortuosity: **a** F0-1Low. 0, 0, and 0 points, **b** H0-1High. 2, 0, and 0 points, **c** F2. 2, 1, and 0 points, **d** F3. 2, 1, and 1 points, **e** F4. 2, 2, and 2 points, respectively

present in 31.9% of patients. Detailed demographic information on the different stages of fibrosis is presented in Additional file 1.

Differences in accuracy by number of classifications in AI-judged stacked microvascular images

The patients with CLD were divided into five groups by the fibrosis stage: F0-1Low (< 5.0 kPa), F0−1High (≥ 5.0 kPa), F2, F3, and F4. The diagnostic accuracy of each classification model was compared when the five different fibrosis stages were classified into five (F0−1Low, F0−1High, F2, F3, and F4), three (F0−1Low, F0−1High, and F2−4), and two (F0−1 and F2−4) classes. Accuracy was 53.8% for the 5-class classification, 66.3% for the 3-class classification, and 83.8% for the 2-class classification, with accuracy increasing as the number of classifications decreased (Table 2). Similar trends were observed for precision, recall, and F1 scores. Comparing human judgement with AI judgement, the AI analysis was a superior model with a higher F1 score in the 2-class classification. In the 3- and 5-class classifications, the performance of both AI and human judgments was almost equal. The confusion matrix for each classification model is shown in Fig. 5.

Determination of optimal cut-off value in a human-judged SMVI scoring system

The SMVI scoring system is an assessment method in which the examiner makes a subjective judgment based on three vascular morphological hallmarks (narrowing, caliber irregularity, and tortuosity). ROC analysis was performed to determine the SMVI score that could distinguish significant fibrosis (≥ F2). The optimal cut-off value was 3 points, yielding a sensitivity of 82.4% and specificity of 90.3% (AUC = 0.93, 95% confidence interval: 0.91−0.95).

Mild fibrosis versus significant fibrosis classifier performance metrics

The diagnostic accuracy of AI-judged and human-judged classifications was compared. AI showed superior accuracy in diagnosing fibrosis stages ≥ F2, with an accuracy rate of 83.8% for the AI versus 81.6% for visual evaluation by a human examiner (t-test, $p < 0.05$). Similarly, AI demonstrated higher sensitivity and specificity of 84.2% and 83.5%, respectively. The details of the classification performance are listed in Table 3.

Representative image of SMVI determined by AI

Examples of images correctly classified into fibrosis stages using the CNN model are shown in Fig. 5. In F0−1 cases, many images showed thick vessels, smooth contours, and linear runs, whereas in F2−4 cases, the vessels were thin, contours were irregular, and runs were nonlinear. An example of an incorrectly classified stacked microvascular image is shown in Fig. 6. Although it was not possible to identify classification regularities common to all cases, images of vessels with thick, well-defined contours were classified as F0−1, whereas those with narrow, complex running vessels were classified as F2−4.

Table 1 Demographic information for all participants

Parameter	Participants (*n* = 517)
Age (years)	63.6 ± 12.9
Sex (male/female)	254/263
Etiology, n (%)	
HCV	167 (32.3)
HBV	126 (24.3)
MASLD	90 (17.4)
MetALD	7 (1.4)
Cryptogenic SLD	4 (0.8)
ALD	78 (15.1)
AIH	14 (2.7)
PBC	12 (2.3)
Drug	5 (1.0)
Unknown	14 (2.7)
LSM-based fibrosis stage, n (%)	
F0–1Low	243 (47.0)
F0–1High	109 (21.1)
F2	43 (8.3)
F3	49 (9.5)
F4	73 (14.1)
SMVI score, n (%)	
0	140 (27.1)
1	99 (19.1)
2	108 (20.9)
3	64 (12.4)
4	56 (10.8)
5	28 (5.4)
6	22 (4.3)

Values are presented as mean ± standard deviation or number (%). All participants are Japanese patients

HCV Hepatic C virus, *HBV* Hepatic B virus, *MASLD* Metabolic dysfunction associated with steatotic liver disease, *ALD* Alcohol-associated liver disease, *AIH* Autoimmune hepatitis, *LSM* Liver stiffness measurement, *PBC* Primary biliary cholangitis

Discussion

Our study revealed the usefulness of deep learning as an analytical method for stacked microvascular images. In particular, we showed that AI-based SMVI analysis is useful for detecting significant fibrosis ($\geq$ F2) in patients with CLD.

CLD, characterized by inflammation and progressive fibrosis, has been shown to undergo hepatic angiogenesis regardless of etiology [22]. In cirrhosis, the enlarged pseudolobular nodules compress the portal and hepatic venous branches, resulting in narrowing and tortuosity of the intrahepatic vessels [23–26]. In recent years, such modifications of vessel architecture have been analyzed using high-sensitivity Doppler methods, such as Superb microvascular imaging and microvascular imaging. However, because these modalities yield images as a single two-dimensional section, the vessels depicted are often fragmented, and the results are difficult to interpret [27, 28]. As a solution to these problems, an ultrasound Doppler technique called SMVI, which provides enhanced images by image accumulation, was developed, which allowed detailed visualization of intrahepatic blood vessels [9]. Furthermore, a scoring method based on three hallmarks was devised to analyze stacked microvascular images, but the possibility of potential subjective bias could not be ruled out because the system was dependent on human judgment. Therefore, we attempted to judge the stacked microvascular images here using AI, without depending on the human eye.

By introducing a residual learning approach, ResNet50 revolutionized the way deep networks are trained, enabling the development of deeper networks without sacrificing performance or stability. The ability to efficiently learn very deep architectures and achieve high accuracy in tasks such as image classification has made ResNet50 one of the most influential

Table 2 Differences in the performance by the number of classifications

a)				
Classification	Accuracy↑	Precision↑	Recall↑	F1 score↑
2 class (F0–1, F2–4)	0.838	0.812	0.839	0.821
3 class (F0–1Low, F0–1High, F2–4)	0.663	0.590	0.592	0.585
5 class (F0–1Low, F0–1High, F2, F3, F4)	0.538	0.368	0.395	0.372
b)				
Classification	Accuracy↑	Precision↑	Recall↑	F1 score↑
2 class (F0–1, F2–4)	0.816	0.636	0.775	0.698
3 class (F0–1Low, F0–1High, F2–4)	0.655	0.580	0.586	0.583
5 class (F0–1Low, F0–1High, F2, F3, F4)	0.557	0.375	0.396	0.385

Arrows in the column headers indicate that higher values are superior

a) artificial intelligence-based judgments

b) human examiner-based judgments

Miura *et al. BMC Medical Imaging*   (2025) 25:8

Page 8 of 11
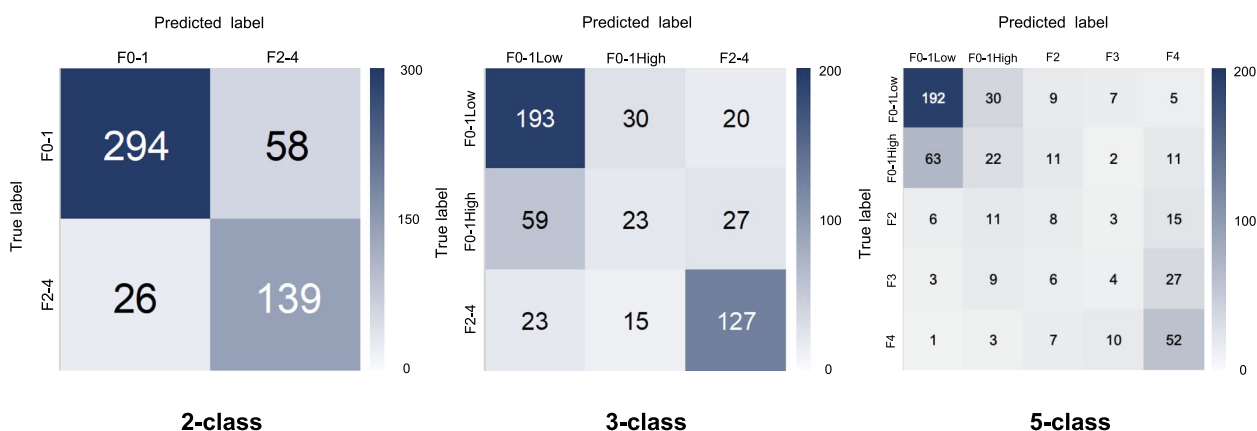
Fig. 5 Confusion matrix showing performance by number of classifications. Confusion matrix of true (row) vs. model-generated predicted (column) classes for fibrosis stage in stacked microvascular image analysis. Values are shown as numbers with shaded colors

Table 3 Comparison of diagnostic performance between human and artificial intelligence-based judgments in 2-class comparisons

	Judge	
	AI (95% CI)	Sonographer (95% CI)
Sensitivity ↑	0.842 (0.778–0.894)	0.775 (0.697–0.840)
Specificity ↑	0.835 (0.792–0.872)	0.832 (0.790–0.868)
PPV ↑	0.706 (0.637–0.768)	0.636 (0.559–0.708)
NPV ↑	0.919 (0.883–0.946)	0.907 (0.871–0.936)
Accuracy ↑	0.838 (0.803–0.868)	0.816 (0.780–0.849)
LR+ ↑	5.113 (4.005–6.527)	4.611 (3.620–5.874)
LR- ↓	0.189 (0.132–0.269)	0.271 (0.199–0.369)

Values in Table 3 are expressed as point estimates and 95% CI. The up arrow in the first column indicates that a higher value is superior, and the down arrow indicates that a lower value is superior

Abbreviations: *AI* Artificial intelligence, *CI* Confidence interval, *PPV* Positive-predictive value, *NPV* Negative-predictive value, *LR +* Positive-likelihood ratio, *LR-* Negative-likelihood ratio

architectures in deep learning. Therefore, ResNet50 is a more advanced deep learning model that adds a new approach to the traditional CNN architecture, but recent AI models have evolved significantly. Zhang et al. reported that region-based integration-and-recalibration networks [29], regional context-based recalibration network [30], and pyramid pixel context adaption modules [31] help to improve medical image classification performance, demonstrating their superiority over recent deep neural networks. In this study, we used a basic ResNet50 architecture as an initial step of building an AI model for monitoring fibrosis, but by incorporating the recent architectures listed above, our method can enhance the feature extraction ability from medical images and improve its performance.

Vision Transformer (ViT) is a type of deep learning model that uses transformer architecture for image recognition tasks. Specifically, it uses an approach that is different from traditional CNNs by utilizing a transformer architecture. Transformers are originally models often used in natural language processing, mainly because of their ability to focus on important parts of the data by leveraging self-attention mechanisms. In the field of deep learning, CNNs have been the dominant method for tasks such as image recognition and object detection, but ViT has been touted as an alternative. While they perform strongly, especially on large datasets, their effectiveness may be limited on small datasets or with constrained computational resources. In contrast, CNNs are good at capturing local features of images, process images using convolutional layers, and generally have the advantage of capturing fine features of images (edges, textures, etc.). In addition to the ResNet50 we used in this study, leveraging newer computer vision classification models such as ViT [32], Swin Transformer [33], RepViT [34], and SLaK [35] in the future may be useful.

The AI judgments showed a decrease in accuracy when the number of classes increased: accuracy was 83.8% when using two classes, 66.3% when using three classes, and 53.8% when using five classes for classification. This may be due to the imbalance in sample size between each fibrosis stage and the small total sample size. At present, the diagnostic performance of the 3- and 5-class classifications is not good but may improve with larger sample sizes. On the other hand, the 2-class classification has excellent diagnostic performance, and the AI analysis of SMVI has a significant role as a screening tool to effectively detect significant fibrosis. Furthermore, especially in the 2-class classification, the F1 score was 0.821 for the

Miura *et al. BMC Medical Imaging*   (2025) 25:8

Page 9 of 11

Predicted label
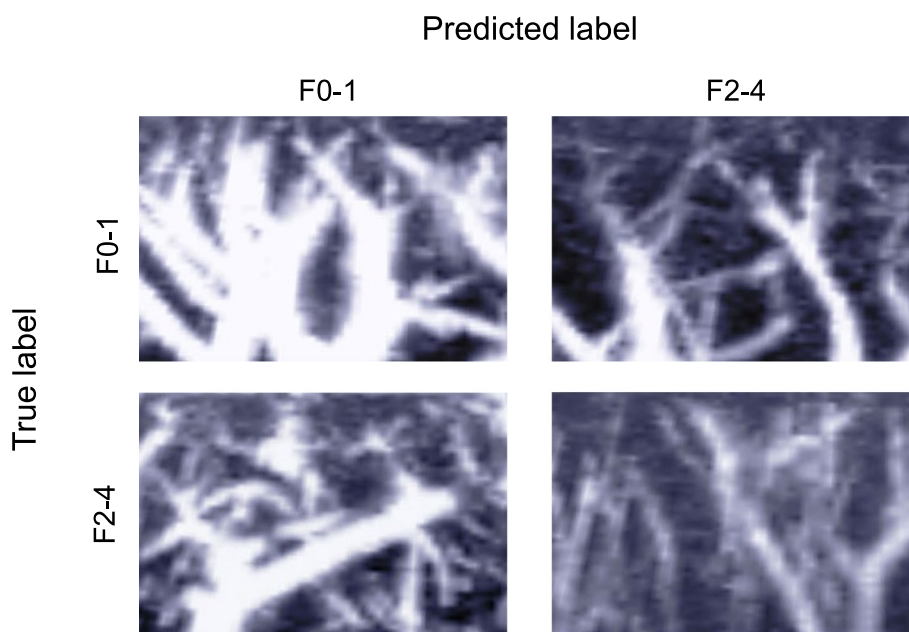
Fig. 6 Representative images of stacked microvascular imaging by convolutional neural network models for fibrosis stage prediction. The size of the cropped image is 200×300 pixels (22×14 mm)

AI analysis, which is higher than that for the human judgment, indicating that the AI analysis is better at fitting the model.

Since significant fibrosis (≥F2) is a risk factor for cirrhosis and overall mortality [36], this study focused on a 2-class classification for the detection of significant fibrosis (≥F2). The accuracy of diagnosis in the significant fibrosis group was 83.8% for AI and 81.6% for human judgment. The sensitivity was 84.2% for AI and 77.5% for human judgment, and the specificity was 83.5% for AI and 83.2% for human judgement, with the AI analysis performing slightly better in both (Table 3). Therefore, when determining significant fibrosis progression, the diagnostic accuracy being higher with AI analysis than with conventional SMVI scoring methods is a considerable improvement. Importantly, the use of AI allows objective decisions to be made without the possibility of subjective bias being introduced by human examiners. Improved objectivity in ultrasonography will lead to a reduction in the interrater differences attributed to differences in experience and technical skills. The development of AI in the medical field is accelerating. Expectations are particularly high for "AI-computer-aided detection" (AI-CAD)," which combines computer-aided diagnosis and AI [37].

The SMVI scoring system is particularly sensitive to early fibrosis, as compared to conventional fibrosis assessment methods, because the SMVI scoring system is more likely to detect vascular narrowing, which is an early change in liver fibrosis [9]. On the other hand, AI analysis of SMVI was effective in identifying fibrosis progression groups. This may be because the scoring system is a systematic assessment based on the individual vessel characteristics of narrowing, caliber irregularity, and tortuosity, whereas AI analysis identifies image patterns non-systematically. In future, SMVI may be widely implemented in daily clinical practice as a liver fibrosis evaluation method if general-purpose ultrasound systems are equipped with SMVI as an AI-CAD.

The SMVI technique is considered less susceptible to the effects of hepatic congestion and acute inflammation than elastography because the analysis is based on the morphological running of the blood vessels. Therefore, elastography and SMVI have complementary roles, and their combination may improve the diagnostic performance of the liver fibrosis stage. By determining early fibrosis with conventional SMVI scoring methods and objectively containing the fibrosis progression group with this AI analysis, SMVI can be considered a validated technique to compensate for the weakness of elastography.

Although AI judgments slightly outperformed human judgments in the 2-class classification in this study, it is still difficult to conclude that the current model adequately meets the needs of clinical diagnostic support. At present, we believe that the best clinical diagnosis support model would be to use AI to identify F2–4 and to automatically measure mean vessel diameter (narrowing)

Miura *et al. BMC Medical Imaging*   (2025) 25:8

Page 10 of 11

for F0–1. Recent progress in AI models has been remarkable, and the latest deep learning model can possibly be used to improve the 5-class diagnostic performance of the AI-only model.

This study has some limitations. First, the fibrosis stage used in this study was based on elastography, rather than on liver biopsy. However, it has been reported that the need for liver biopsy is only 3% when the EASL algorithm is applied in a primary care/diabetes clinic cohort [38]. Moreover, in clinical practice, non-invasive tests using LSM values are widely used to diagnose liver fibrosis stages, limiting the need for highly invasive liver biopsies. Second, this study involved a small sample size (517 cases) for a deep learning study. Thus, for future research, we aim to conduct a multicenter, prospective study to evaluate the fibrosis diagnostic performance analysis based on AI analysis using the SMVI method.

Conclusion

In conclusion, the SMVI method, known for its high sensitivity Doppler accumulation, has demonstrated enhanced objectivity with the integration of AI technology. It has proven to be particularly effective in assessing fibrosis progression beyond the F2 stage in patients with CLD.

Abbreviations

AI	Artificial intelligence
CLD	Chronic liver disease
SMVI	Stacked microvascular imaging
EASL	European Association for the Study of the Liver
LSM	Liver stiffness measurement
VCTE	Vibration-controlled transient elastography
CNN	Convolutional neural networks
ROC	Receiver operating characteristic
AUC	Area under the curve
AI-CAD	Artificial intelligence-computer-aided detection
ViT	Vision transformer

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-024-01531-x.

Supplementary Material 1.

Declarations

Ethics approval and consent to participate
This study was reviewed and approved by the Fukuoka Tokushukai Hospital Institutional Review Board (Approval Number, 220101), and informed consent was obtained from all patients included in the study. All study activities were performed in accordance with relevant guidelines and regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Ultrasound and Clinical Laboratory, Fukuoka Tokushukai Hospital, Fukuoka 816-0864, Japan. [2]Department of Laboratory Science, Yamaguchi University Graduate School of Medicine, Yamaguchi 755-8508, Japan. [3]Department of Information Science and Engineering, Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi 755-8611, Japan.

References

1. Roehlen N, Crouchet E, Baumert TF. Liver fibrosis: mechanistic concepts and therapeutic perspectives. Cells. 2020;9:875.
2. Staufer K, Stauber RE. Steatotic liver disease: metabolic dysfunction, alcohol, or both? Biomedicines. 2023;11:2108.
3. Ghany MG, Strader DB, Thomas DL, Seeff LB, American Association for the study of liver diseases. Diagnosis, management, and treatment of hepatitis C: an update. Hepatology. 2009;49:1335–74.
4. Thampanitchawong P, Piratvisuth T. Liver biopsy: complications and risk factors. World J Gastroenterol. 1999;5:301–4.
5. Yang J, Li J, Ye G, Luo Y. Comparison of visual transient elastography and shear wave elastography in evaluating liver fibrosis in patients with chronic liver disease. Int J Gen Med. 2021;14:3553–61.
6. Kovatsch A, Honcharova-Biletska H, Segna D, Steigmiller K, Blümel S, et al. Performance of two-dimensional shear wave elastography and transient elastography compared to liver biopsy for staging of liver fibrosis. Eur J Clin Investig. 2023;53:1–10.
7. European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu, Clinical Practice Guideline Panel, Chair:, EASL Governing Board representative:, & Panel members. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol. 2021;75:659-89.
8. Barr RG, Wilson SR, Rubens D, Garcia-Tsao G, Ferraioli G. Update to the society of radiologists in ultrasound liver elastography consensus statement. Radiology. 2020;296:263–74.
9. Miura D, Suenaga H, Ichihara K. The utility of a novel stacked microvascular imaging for enhanced detection of fibrosis in chronic liver diseases. Ultrasound Med Biol. 2024;50:975–84.
10. Imajo K, Honda Y, Kobayashi T, Nagai K, Ozaki A, Iwaki M, et al. Direct comparison of US and MR elastography for staging liver fibrosis in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol. 2022;20:908–17.
11. Ma J, Wu F, Jiang T, Zhu J, Kong D. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. Med Phys. 2017;44:1678–91.
12. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. J Digit Imaging. 2017;30:477–86.

13. Fujioka T, Mori M, Kubota K, Oyama J, Yamaga E, Yashima Y, et al. The utility of deep learning in breast ultrasonic imaging: a review. Diagnostics (Basel). 2020;10:1055.

14. Wang B, Perronne L, Burke C, Adler RS. Artificial intelligence for classification of soft-tissue masses at US. Radiol Artif Intell. 2021;3:e200125.

15. Nakata N, Siina T. Ensemble learning of multiple models using deep learning for multiclass classification of ultrasound images of hepatic masses. Bioengineering (Basel). 2023;10:69.

16. Park HC, Joo Y, Lee OJ, Lee K, Song TK, Choi C, et al. Automated classification of liver fibrosis stages using ultrasound imaging. BMC Med Imaging. 2024;24:36.

17. Yin Y, Yakar D, Dierckx RAJO, Mouridsen KB, Kwee TC, de Haas RJ. Liver fibrosis staging by deep learning: a visual-based explanation of diagnostic decisions of the model. Eur Radiol. 2021;31:9620–7.

18. Zhu Z, Lv D, Zhang X, Wang SH, Zhu G. Deep learning in the classification of stage of liver fibrosis in chronic hepatitis B with magnetic resonance ADC images. Contrast Media Mol Imaging. 2021;2021:2015780.

19. Maira D, Cassinerio E, Marcon A, Mancarella M, Fraquelli M, Pedrotti P, et al. Progression of liver fibrosis can be controlled by adequate chelation in transfusion-dependent thalassemia (TDT). Ann Hematol. 2017;96:1931–6.

20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;770–8. https://doi.org/10.1109/CVPR.2016.90.

21. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant. 2013;48:452–8.

22. Aydın MM, Akçalı KC. Liver fibrosis. Turk J Gastroenterol. 2018;29:14–21.

23. Yamamoto T, Kobayashi T, Phillips MJ. Perinodular arteriolar plexus in liver cirrhosis. Scanning electron microscopy of microvascular casts. Liver. 1984;4:50–4.

24. Baus WP, Hoffbauer FW. Vascular changes in the cirrhotic liver as studied by the injection technic. Digest Dis Sci. 1963;8:689–700.

25. Hano H, Takasaki S. Three-dimensional observations on the alterations of lobular architecture in chronic hepatitis with special reference to its angioarchitecture for a better understanding of the formal pathogenesis of liver cirrhosis. Virchows Arch. 2003;443:655–63.

26. Shibayama Y, Nakata K. The role of pericentral fibrosis in experimental portal hypertension in rats. Liver. 1991;11:94–9.

27. Kuroda H, Abe T, Kakisaka K, Fujiwara Y, Yoshida Y, Miyasaka A, et al. Visualizing the hepatic vascular architecture using superb microvascular imaging in patients with hepatitis C virus: A novel technique. World J Gastroenterol. 2016;22:6057–64.

28. Tosun M, Uslu H. Comparison of superb microvascular imaging and shear wave elastography for assessing liver fibrosis in chronic hepatitis B. Ultrasonography. 2022;41:394–402.

29. Zhang X, Xiao Z, Fu H, Hu Y, Yuan J, Xu Y, et al. Attention to region: Region-based integration-and-recalibration networks for nuclear cataract classification using AS-OCT images. Med Image Anal. 2022;80:102499.

30. Zhang X, Xiao Z, Yang B, Wu X, Higashita R, Liu J. Regional context-based recalibration network for cataract recognition in AS-OCT. Pattern Recognit. 2024;147:110069.

31. Zhang X, Xiao Z, Wu X, Chen Y, Zhao J, Hu Y, et al. Pyramid pixel context adaption network for medical image classification with supervised contrastive learning. IEEE Trans Neural Netw Learn Syst. 2024. https://doi.org/10.1109/TNNLS.2024.3399164.

32. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. ArXiv. 2022. https://doi.org/10.48550/arXiv.2010.11929.

33. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision. ArXiv. 2021:10012–22. https://doi.org/10.1109/ICCV48922.2021.00986.

34. Wang A, Chen H, Lin Z, Han J, Ding G. RepViT: Revisiting mobile CNN from ViT perspective. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ArXiv. 2024:15909-20. https://doi.org/10.48550/arXiv.2307.09283.

35. Liu S, Chen T, Chen X, Chen X, Xiao O, Wu B, et al. More ConvNets in the 2020s: Scaling up Kernels Beyond 51x51 using Sparsity. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ArXiv. 2022. https://doi.org/10.48550/arXiv.2207.03620.

36. Lomonaco R, Godinez Leiva E, Bril F, Shrestha S, Mansour L, et al. Advanced liver fibrosis is common in patients with type 2 diabetes followed in the outpatient setting: the need for systematic screening. Diabetes Care. 2021;44:399–406.

37. Fujita H. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. Radiol Phys Technol. 2020;13:6–19.

38. Canivet CM, Costentin C, Irvine KM, Delamarre A, Lannes A, Sturm N, et al. Validation of the new 2021 EASL algorithm for the noninvasive diagnosis of advanced fibrosis in NAFLD. Hepatology. 2023;77:920–30.

Publisher's Note