

Diagnosis of Significant Liver Fibrosis in Patients With Chronic Hepatitis B By Using a Deep Learning-Based Data Integration Network

Zhong Liu

Shenzhen University <https://orcid.org/0000-0001-9650-6097>

Huiying Wen

Shenzhen University

Ziqi Zhu

Shenzhen University

Qinyuan Li

Affiliated Hospital of Guangdong Medical University

Li Liu

The Chinese University of Hong Kong

Tianjiao Li

Shenzhen University

Wencong Xu

Shenzhen University

Chao Hou

Shenzhen University

Bin Huang

Shenzhen University

Zhiyan Li

Shenzhen Third People's Hospital

Changfeng Dong (✉ chaosheng-02@szsy.sustech.edu.cn)

Shenzhen Third People's Hospital

Xin Chen

Shenzhen University <https://orcid.org/0000-0002-5406-4136>

Research Article

Keywords: Fatty liver disease, Liver fibrosis, Significant liver fibrosis, Chronic hepatitis B, Ultrasonography, Liver elastography, Deep learning, Feature fusion, Non-invasive fibrosis staging, Fibrosis biomarker

Posted Date: November 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1052634/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Hepatology International on March 21st, 2022. See the published version at <https://doi.org/10.1007/s12072-021-10294-4>.

Abstract

Background and aims:

Chronic hepatitis B virus (CHB) infection remains a major global health burden and the non-invasive and accurate diagnosis of significant liver fibrosis (\geq F2) in CHB patients is clinically very important. This study aimed to assess the potential of the joint use of ultrasound images of liver parenchyma, liver stiffness values, and patients' clinical parameters in a deep learning model to improve the diagnosis of \geq F2 in CHB patients.

Methods

Of 527 CHB patients who underwent US examination, liver elastography and biopsy, 284 eligible patients were included. We developed a deep learning-based data integration network (DI-Net) to fuse the information of ultrasound images of liver parenchyma, liver stiffness values and patients' clinical parameters for diagnosing \geq F2 in CHB patients. The performance of DI-Net was cross-validated in a main cohort (n =155) of the included patients and externally validated in an independent cohort (n = 129), with comparisons against single-source data-based models and other non-invasive methods in terms of the area under the receiver-operating-characteristic curve (AUC).

Results

DI-Net achieved an AUC of 0.943 (95% confidence interval [CI]: 0.893-0.973) in the cross-validation, and an AUC of 0.901 (95% CI: 0.834-0.945) in the external validation, which were significantly greater than those of the comparative methods (AUC ranges: 0.774-0.877 and 0.741-0.848 for cross- and external validations, respectively, $P_s < 0.01$).

Conclusion

The joint use of ultrasound images of liver parenchyma, liver stiffness values, and patients' clinical parameters in a deep learning model could significantly improve the diagnosis of \geq F2 in CHB patients.

Introduction

Chronic hepatitis B virus (CHB) infection remains a major global health burden, with approximately 257 million people infected worldwide and 800,000 deaths annually¹⁻³. Liver structure and function alternations in CHB patients may lead to the development of liver fibrosis. According to the METAVIR system⁴, liver fibrosis can be stratified into the following stages: F0, no fibrosis; F1, portal fibrosis without septa; F2, portal fibrosis and few septa; F3, numerous septa without cirrhosis; F4, cirrhosis. Without

proper treatment, liver fibrosis may develop from stages of F0-F4 to hepatocellular carcinoma (HCC) or even liver failure^{2,3,5}, however, if treated in time, the ongoing progression of fibrosis can be reversed³. Therefore, for better management of liver fibrosis in CHB patients, an accurate diagnosis of the fibrosis stage is very important.

Currently, liver biopsy remains the gold-standard for liver fibrosis staging, but it is associated with potential complications and subject to several limitations, such as sampling error and inter-observer variability. As an alternative to liver biopsy, image analysis based on computed tomography, magnetic resonance or ultrasound (US) imaging is usually used for non-invasive staging of liver fibrosis. Among the various imaging modalities, US is preferred due to its non-ionizing radiation and wide availability. Studies have shown that US image features, such as an uneven or undulating liver surface, heterogeneous echo texture of the liver parenchyma; and changes in the diameters of vessels, blood flow velocity, and spleen size are correlated with liver fibrosis⁶⁻⁸. Hence, in clinical practice visual assessment of the US image features is often performed for fibrosis staging and screening prior to liver biopsy. However, visual assessment on US images is subjective and its accuracy quite depends on the experience of radiologists.

Apart from visual assessment, there also exist other non-invasive methods for liver fibrosis staging, such as the fibrosis biomarkers based on patients' clinical parameters⁹⁻¹¹, fibrosis detectors based on liver stiffness measurements (LSMs)^{12,13}, and artificial intelligence (AI) based on US images¹⁴⁻¹⁶. Particularly, deep convolutional neural network (DCNN), a subtype of AI, has been rapidly developed and is becoming a promising tool for US image analysis. By using supervised learning based on a large dataset of labeled images, a DCNN model can be trained and used for an objective and intelligent evaluation of the disease reflected by US images. To date, many studies have been conducted to explore the potential of DCNN models for liver fibrosis assessment in US images. However, the already-developed DCNN models often performed well in diagnosing liver cirrhosis (F4), but exhibited poor performance in the detection of significant liver fibrosis (\geq F2)^{14,15}, while the identification of \geq F2 is particularly important as it signals the necessity for anti-fibrotic treatments¹⁷. Thus, it is imperative to develop a new DCNN model for better diagnosis of \geq F2.

Noting that most of the already-developed deep learning models are based on the mere use of single-source data of US images, neglecting other data sources, such as the previously mentioned LSMs and the patients' clinical parameters which also have potential to reflect the liver fibrosis stages, it is meaningful to investigate if the performance of DCNN models in diagnosing \geq F2 could be further enhanced by using multi-source data. Thus, it was the aim of this study to assess the potential of the joint use of US images of liver parenchyma, liver stiffness values, and patients' clinical parameters in a deep learning model to improve the diagnosis of \geq F2 in CHB patients.

Patients And Methods

Patients

All consecutive patients with CHB (hepatitis B surface antigen positive for more than 6 months), who underwent liver biopsy from May 2016 to January 2021 at Shenzhen Third People's Hospital (SZTPH), Shenzhen, China, were studied. The patients who had received US B-mode examination, and LSM via point-shear-wave elastography (P-SWE) within one week prior to liver biopsy were included. Exclusion criteria were: (a) age less than 18 years; (b) received antiviral treatment in 6 months prior to liver biopsy; (c) combined with any other liver diseases, including autoimmune hepatitis, alcoholic liver disease, ascites, and HCC; (d) co-infection with any other viral hepatitis; (e) liver samples smaller than 10 mm or containing less than 6 portal tracts; (f) missing required serological results; (g) unsuccessful P-SWE. According to the eligibility criteria, we successfully enrolled 385 patients, among which 284 patients who had complete data required were finally included for analysis (Figure 1). This study was approved by the ethics committee of SZTPH. Written informed consent was obtained from each enrolled patient.

US B-mode and P-SWE Examinations

A conventional US examination followed by a P-SWE was conducted after the patient had undergone an overnight fast (at least 8 hours). The examinations were performed by using one of the following two US systems: Philips iU22, and Mindray Resona7. Both systems have the function of P-SWE and are routinely used in SZTPH. For the conventional US examination, a linear-array probe (L9-3 for Philips iU22, L11-3U for Mindray Resona7) was used to capture US B-mode images of the liver surface, while a convex-array probe (C14L5 for Philips iU22, SC6U-1 for Mindray Resona7) was used to capture B-mode images of the liver parenchyma and spleen. During the examination, the patient was instructed to maintain the supine position with both hands extended to the head, and static B-mode images were scanned. Right after the conventional US examination, the patient was required to maintain the supine position with right arm at maximum abduction. Then, the sonographer changed the scanning mode to P-SWE for LSMs. Among the included 284 patients, 155 received examinations on the Philips system, and the rest (n = 129) received examinations on the Mindray system.

Clinical Data Acquisition

The clinical data used in our study include patient demographic and serological parameters, which were determined on the same day of US examinations. The demographic data contain five parameters, including sex, age, height, weight, and body mass index. The serological data comprise eighteen parameters^{18,19}, as follows: peripheral platelet count (PLT), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyltransferase (GGT) level, serum albumin (ALB), alkaline phosphatase (ALP), glucose (GLU), total cholesterol (TC), total bilirubin (TB), triacylglycerol, uric acid (UA), C-reactive protein (CRP), and serous markers of hyaluronic acid (HA), amino terminal of procollagen type III peptide (PIIIP), collagen type IV (CIV), laminin (LN), AST/PLT, and AST/ALT.

Pathological Examination

The pathological examination was performed within less than one week following the US examination. A US-guided percutaneous liver biopsy on the right liver lobe was performed by using a 16-gauge biopsy

needle (MC1616 [16G, 16cm], BARD, New Jersey, USA). The biopsied liver sample was 10-20 mm in length, containing at least 6 portal areas, which was made into paraffin sections and stained with Sirius Red (S1020, Wuhan Haotian Bioscience Technology Limited Company, Hubei, China). Liver fibrosis was semi-quantitatively assessed according to the METAVIR system⁴, and grades \geq F2 indicated significant fibrosis. The pathological results served as ground truth for the subsequent model training and validation.

Model Development

We devised a deep learning-based data integration network (DI-Net) to fuse the information of US images of liver parenchyma, liver stiffness values and patients' clinical parameters for diagnosing \geq F2 in the included CHB patients. The DI-Net (Figure 2) has two branches: Branch-1 and Branch-2, respectively with input of B-mode image of liver parenchyma and clinical + LSM data. Branch-1 is a DCNN model backbone by ResNet50²⁰ for automatic extraction of deep features from B-mode images (the feature set is termed as B, 512 in size, see details in the Supplementary Material). Branch-2 compacted the clinical parameters (C) and liver stiffness values (S) into a single feature vector (C+S, 23 in size). At the backend of the model (Figure 2, Step-3), the two branch features were fused via an efficient feature fusion module which consists of three feature units (FUs). The first FU, FU-1, has two fully-connected (FC) layers with the number of neurons reduced from 512 to 23, such that the B features stemming from Branch-1 can be condensed to a feature set with a dimension matched with the number of C+S parameters. The second FU, FU-2, also has two FC layers, and the number of neurons for each layer is identical to that of the C+S parameters. As such, the C+S parameters can be combined internally first to enhance their associations with the class labels prior to the fusion with the B features. Finally, the feature vectors output from FU-1 and FU-2, each with a size of 23, were concatenated together by FU-3 to yield a comprehensive feature set (46 in size), which was then processed by two FC layers (the number of neurons are 46 and 2, respectively, see Step-4 of Figure 2) for predicting the ultimate class labels: \geq F2 or $<$ F2.

Model Training and Validation

Patients were divided into two cohorts according to the US systems under which they received examinations. The patients who underwent examinations on the Philips system (n = 155) contributed to a main cohort, and the corresponding data yielded an internal dataset for model training and cross-validation. The remaining patients who underwent examinations on the Mindray system (n =129) served as an independent cohort, and the corresponding data yielded an additional dataset for external validation of the model trained on the internal dataset.

We used five-fold cross-validation to train and validate the DI-Net on the internal dataset. Specifically, the data in the internal dataset was divided into five subsets with no patient overlap, among which three subsets was used for training (see details in the Supplementary Material), one of the remaining two subsets for model optimization, and the rest for validation. By altering the training, optimization and validation folds five times, each fold was tested once. The models trained in the internal dataset were

additionally validated on the external dataset. As there were five models trained on the internal dataset through cross-validation, the class probabilities predicted by the models were averaged for final decision making in the external validation.

Evaluation Metrics and Comparative Methods

The diagnostic capacity of DI-Net was evaluated in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), the receiver-operating-characteristic (ROC) curve, and the area under the ROC curve (AUC). We compared the performance of DI-Net against the following methods: 1) single-source data-based models, i.e. the models trained with single-source data of B-mode images, liver stiffness values, or patients' clinical parameters. 2) fibrosis biomarkers^{9,10}, including APRI and FIB-4. 3) interpretations by skilled radiologists. More details of the comparative methods are presented in the Supplementary Material.

Statistical Analysis

Continuous variables of the patients' data were expressed as mean \pm standard deviation and compared by Student's *t* test, while the categorical variables were expressed as a number (percentage) and compared by chi-square test. For the quantitative results of ACC, SEN, SPE, PPV, NPV and AUC, 95% confidence intervals (CIs) were calculated using the Clopper-Pearson method²¹. We used DeLong test²² to evaluate the differences between the ROC curves and the associated AUC values. *P* values were two-sided, and the values less than 0.05 indicated statistical significance. All statistical tests were performed on the software of MedCalc (Version 18.2) and MATLAB (Version R2019a).

Results

Patient Characteristics

A total of 527 CHB patients who underwent liver biopsy, conventional US and P-SWE examinations were studied. 142 (26.9%) patients did not meet the eligibility criteria and were excluded (Figure 1). Among the enrolled 385 patients, 101 (26.2%) patients were with incomplete data: 73 without complete required serological parameters, and 28 without successful P-SWE measurements. Finally, we included 284 patients for analysis, of which 155 subjects (mean age, 37.9 years, standard deviation, 10.3 years; F0-F1, 68, F2-F4, 87) made up the main cohort and 129 subjects (mean age, 37.7 years, standard deviation, 8.3 years; F0-F1, 73, F2-F4, 56) made up an independent cohort. The main clinical and demographic characteristics of the study cohort are summarized in Table 1.

Cross-validation Results

Through cross-validation on the internal dataset, DI-Net achieved an AUC of 0.943 [95% CI, 0.893-0.973], which was significantly higher than those yielded by the comparative methods, including the models trained by using only B-mode images (AUC 0.873 [95% CI, 0.808-0.919]), liver stiffness values (AUC 0.877

[0.815-0.925]), or clinical data (AUC 0.843 [0.778-0.898]), all P s <0.01, DeLong test. Quantitative results of the AUC, ACC, SEN, SPE, PPV, and NPV of different models are presented in the upper four rows parallel to the cross-validation in Table 2. Noticeably, the quantitative indexes of DI-Net were universally higher than those of any single-source data-based method. The ROC curve of DI-Net is shown in Figure 3a, where the curves of the single-source data-based models are also presented for comparisons. In the same validation, the overall performance of DI-Net was also significantly better than those of APRI (AUC, 0.791 [0.721-0.854]), FIB-4 (AUC, 0.808 [0.737-0.867]) and visual assessments of the two skilled radiologists (AUCs, 0.793 [0.721-0.854] and 0.774 [0.700-0.837]), all P s <0.01, DeLong test. The ROC curves of these non-invasive methods are presented in Figure 3b. The corresponding quantitative results are summarized in the bottom four rows parallel to the cross-validation in Table 2.

External Validation Results

In the external dataset, DI-Net yielded an AUC of 0.901 [95% CI, 0.834-0.945]. Although the AUC value was lower than the one obtained in the internal dataset (0.943 [95% CI, 0.893-0.973]), it was still greater than those achieved by the single-source data-based models: AUCs, 0.806 (0.727-0.870), 0.848 (0.771-0.903), and 0.813 (0.736-0.877) for that trained by using only B-mode images, liver stiffness values, or patients' clinical parameters, respectively, P s <0.01, DeLong test. In the external validation, DI-Net also outperformed APRI, FIB-4 and visual assessments of the two radiologists (AUCs, 0.803 [0.727-0.870], 0.828 [0.753-0.890], 0.753 [0.668-0.824], and 0.741 [0.660-0.817], respectively, P s <0.01, DeLong test). The quantitative results corresponding to the external validation are summarized in Table 2 (see rows parallel to the external validation). The ROC curves yielded by the models with different data sources and those by APRI, FIB-4 and radiologists' interpretations are plotted in Figure 3c and d, respectively.

Discussion

The accurate and non-invasive diagnosis of \geq F2 is crucial for better management of patients suffering from chronic liver disease^{2,3,23}, but achieving this goal is technically challenging. As one type of non-invasive methods, fibrosis biomarkers have the advantages of high applicability, good inter-laboratory reproducibility, and wide availability. However, many studies showed that the biomarkers offered less accuracy in diagnosing significant fibrosis than cirrhosis^{24,25}. In our study, the diagnostic performances of two commonly-used fibrosis biomarkers, APRI and FIB-4, in diagnosing \geq F2 were also unsatisfactory, with AUC values ranging from 0.791 to 0.828. The problem of less accuracy in diagnosing \geq F2 than F4 was also encountered by DCNN models. For instance, in the study conducted by Wang et al.¹⁴, a DCNN model, named DLRE (deep learning radiomics of elastography), was devised for liver fibrosis assessment in US shear wave elastography (SWE) images. On a dataset comprising 1990 SWE images of 398 patients with pathologically-proven fibrosis stages, DLRE yielded an AUC of 0.97 [95% CI 0.94-0.99] for F4, whereas for \geq F2, the obtained AUC was 0.85 [95% CI 0.81-0.89].

The less accuracy of the non-invasive methods in diagnosing \geq F2 than F4 is probably due to the fact that the changes of liver fibrosis at intermediate stages are quite subtle and heterogeneous^{24,26,27}. The

use of single-source data of US images^{14,15}, liver stiffness values¹³, or biomarkers^{11,24,25} may not be sufficient to well capture such changes. The results in the study indicated that the information carried by US B-mode images of liver parenchyma could be complemented by the liver stiffness values and patients' clinical parameters for improving the diagnosis of \geq F2. Therefore, the joint use of multi-source data has the potential to better reflect the subtle changes of liver fibrosis, in comparison to the mere use of single-source data. To further verify this point, we replaced the network in Branch-1 of DI-Net with the other popular DCNN backbones, including VGG16²⁸, Inception V3²⁹, and DenseNet121³⁰, and validated the resulted models on our dataset. The results of the resulted models together with the one based on the backbone model of ResNet50 are summarized in Table 1 of the Supplementary Material. We noted that all models with the additional use of liver stiffness values and patients' clinical parameters achieved significantly better performance than those using only B-mode images, for both cases of cross-validation (AUC ranges: 0.908-0.943 vs. 0.773-0.873, P s <0.01, DeLong test) and external validation (AUC ranges: 0.865-0.901 vs. 0.702-0.806, P s <0.01, DeLong test). These results, to some extent, further increased our credibility of the added value of liver stiffness measurements and clinical parameters to the B-mode images of liver parenchyma for improving the diagnosis of \geq F2.

The performance improvement made by DI-Net also benefitted from the specialized design of the feature fusion module. In this respect, we substituted the feature fusion module in DI-Net with a functional unit that directly concatenated the features output from different branches, while adjusting the number of neurons in the penultimate FC layer accordingly. The new model, named DCNN-DFC (Deep convolutional neural network with direct feature concatenation), was trained and validated by following the same procedure as done for DI-Net. The quantitative results yielded by DCNN-DFC are shown in Table 3. Comparing to the results of the model with B-mode images, the AUC improvement made by DCNN-DFC were 1.7 percentages increase in the cross-validation and 1.9 percentages increase in the external validation, which were lower than those made by DI-Net, i.e. 7.0 percentages increase for the cross-validation and 9.5 percentages increase for the external validation. These results demonstrate that the direct concatenation of two feature sets with highly imbalanced dimension weakens the contribution of the one with smaller number of features, and it is better to concatenate the feature sets after adjusting their dimensions to be at an equivalent level.

Our study has limitations. First, this was a single-centre retrospective study. Acquiring more data from multiple centres is necessary to further validate the robustness of DI-Net. Second, our data excluded patients with non-viral liver diseases, such as the non-alcoholic fatty liver disease which may be a confounding factor for fibrosis analysis.

In conclusion, this study successfully developed a deep learning-based data integration network, namely DI-Net, to integrate the information of US images of liver parenchyma, liver stiffness values and patients' clinical parameters for diagnosing significant liver fibrosis in CHB patients. Both cross- and external validations as well as the comparison of DI-Net against the other non-invasive methods demonstrate that the joint use of the multi-source data in a deep learning model could significantly improve the diagnosis of significant liver fibrosis for CHB patients.

Declarations

Acknowledgements

The authors would like to thank all of the patients for their participation in this study.

Data Availability

Data are available upon reasonable request to the corresponding authors.

Animal Research (Ethics)

Not applicable.

Consent to Participate (Ethics)

This study was approved by the hospital's ethical review board (Shenzhen Third People's Hospital, Shenzhen, China) and fully complied with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all the patients enrolled.

Consent to Publish (Ethics)

All authors approved the publication of the manuscript.

Plant Reproducibility

Not applicable.

Clinical Trial Registration

Not applicable.

Author Contribution

ZL and HW designed the study. ZL, CD and XC had full access to all the data. ZL, HW, ZZ and QL analyzed the data. ZL and HW wrote the manuscript. CD and XC fully supervised the study. All authors provided substantial comments on drafts and approved the final report.

Conflict of Interest

The authors have no conflicts of interest to disclose.

Funding

This work was partially supported by National Natural Science Foundation of China under Grants 81871429 and 61901282.

References

1. Vittal A, Ghany MG. WHO Guidelines for Prevention, Care and Treatment of Individuals Infected with HBV A US Perspective. *Clin Liver Dis*. 2019;23(3):417–432.
2. Lampertico P, Agarwal K, Berg T, et al. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol*. 2017;67(2):370–398.
3. Terrault NA, Lok ASF, McMahon BJ, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology*. 2018;67(4):1560–1599.
4. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology*. 1996;24(2):289–293.
5. Tan M, Bhadoria AS, Cui F, et al. Estimating the proportion of people with chronic hepatitis B virus infection eligible for hepatitis B antiviral treatment worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol*. 2021;6(2):106–119.
6. Zheng R-Q, Wang Q-H, Lu M-D, et al. Liver fibrosis in chronic viral hepatitis: an ultrasonographic study. *World J Gastroenterol*. 2003;9(11):2484.
7. Colli A, Fraquelli M, Andreoletti M, et al. Severe liver fibrosis or cirrhosis: accuracy of US for detection –analysis of 300 cases. *Radiology*. 2003;227(1):89–94.
8. Salvatore V, Borghi A, Peri E, et al. Relationship between hepatic haemodynamics assessed by Doppler ultrasound and liver stiffness. *Dig Liver Dis*. 2012;44(2):154–159.
9. Wai CT, Greenon JK, Fontana RJ, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*. 2003;38(2):518–526.
10. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317–1325.
11. Siddiqui MS, Yamada G, Vuppalanchi R, et al. Diagnostic accuracy of noninvasive fibrosis models to detect change in fibrosis stage. *Clin Gastroenterol Hepatol*. 2019;17(9):1877–1885. e1875.
12. Conti F, Serra C, Vukotic R, et al. Assessment of liver fibrosis with elastography point quantification vs other noninvasive methods. *Clin Gastroenterol Hepatol*. 2019;17(3):510–517. e513.
13. Kakegawa T, Sugimoto K, Kuroda H, et al. Diagnostic Accuracy of Two-Dimensional Shear Wave Elastography for Liver Fibrosis: A Multicenter Prospective Study. *Clin Gastroenterol Hepatol*. 2021.
14. 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.
15. 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.
16. 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 International*. 2021.

17. Petitsclerc L, Sebastiani G, Gilbert G, et al. Liver fibrosis: Review of current imaging and MRI quantification techniques. *Journal of Magnetic Resonance Imaging*. 2017;45(5):1276–1295.
18. Hui AY, Chan HL, Wong VW, et al. Identification of chronic hepatitis B patients without significant liver fibrosis by a simple noninvasive predictive model. *Am J Gastroenterol*. 2005;100(3):616–623.
19. Zeng MD, Lu LG, Mao YM, et al. Prediction of significant fibrosis in HBeAg-positive patients with chronic hepatitis B by a noninvasive model. *Hepatology*. 2005;42(6):1437–1445.
20. He KM, Zhang XY, Ren SQ, et al. Deep Residual Learning for Image Recognition. IEEE Conf Comput Vision Pattern Recognit; 2016; Las Vegas, NV, USA.
21. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26(4):404–413.
22. DeLong ER, DeLong DM, Clarkepearson DI. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845.
23. Lee YA, Wallace MC, Friedman SL. Pathobiology of liver fibrosis: a translational success story. *Gut*. 2015;64(5):830–841.
24. Xiao G, Yang J, Yan L. Comparison of Diagnostic Accuracy of Aspartate Aminotransferase to Platelet Ratio Index and Fibrosis-4 Index for Detecting Liver Fibrosis in Adult Patients With Chronic Hepatitis B Virus Infection: A Systemic Review and Meta-analysis. *Hepatology*. 2015;61(1):292–302.
25. Patel K, Sebastiani G. Limitations of non-invasive tests for assessment of liver fibrosis. *JHEP Rep*. 2020;2(2):100067.
26. Martinez SM, Crespo G, Navasa M, et al. Noninvasive Assessment of Liver Fibrosis. *Hepatology*. 2011;53(1):325–335.
27. Ferraioli G, Wong VW-S, Castera L, et al. Liver Ultrasound Elastography: An Update to the World Federation for Ultrasound in Medicine and Biology Guidelines and Recommendations. *Ultrasound Med Biol*. 2018;44(12):2419–2440.
28. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Int Conf Learn Represent; 2015; San Diego, CA, US.
29. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision. IEEE Conf Comput Vision Pattern Recognit; 2016; Las Vegas, NV, USA.
30. Huang G, Liu Z, Van der Maaten L, et al. Densely Connected Convolutional Networks. IEEE Conf Comput Vision Pattern Recognit; 2017; Honolulu, HI, USA.

Tables

Due to technical limitations, table 1,2,3 is only available as a download in the Supplemental Files section.

Figures

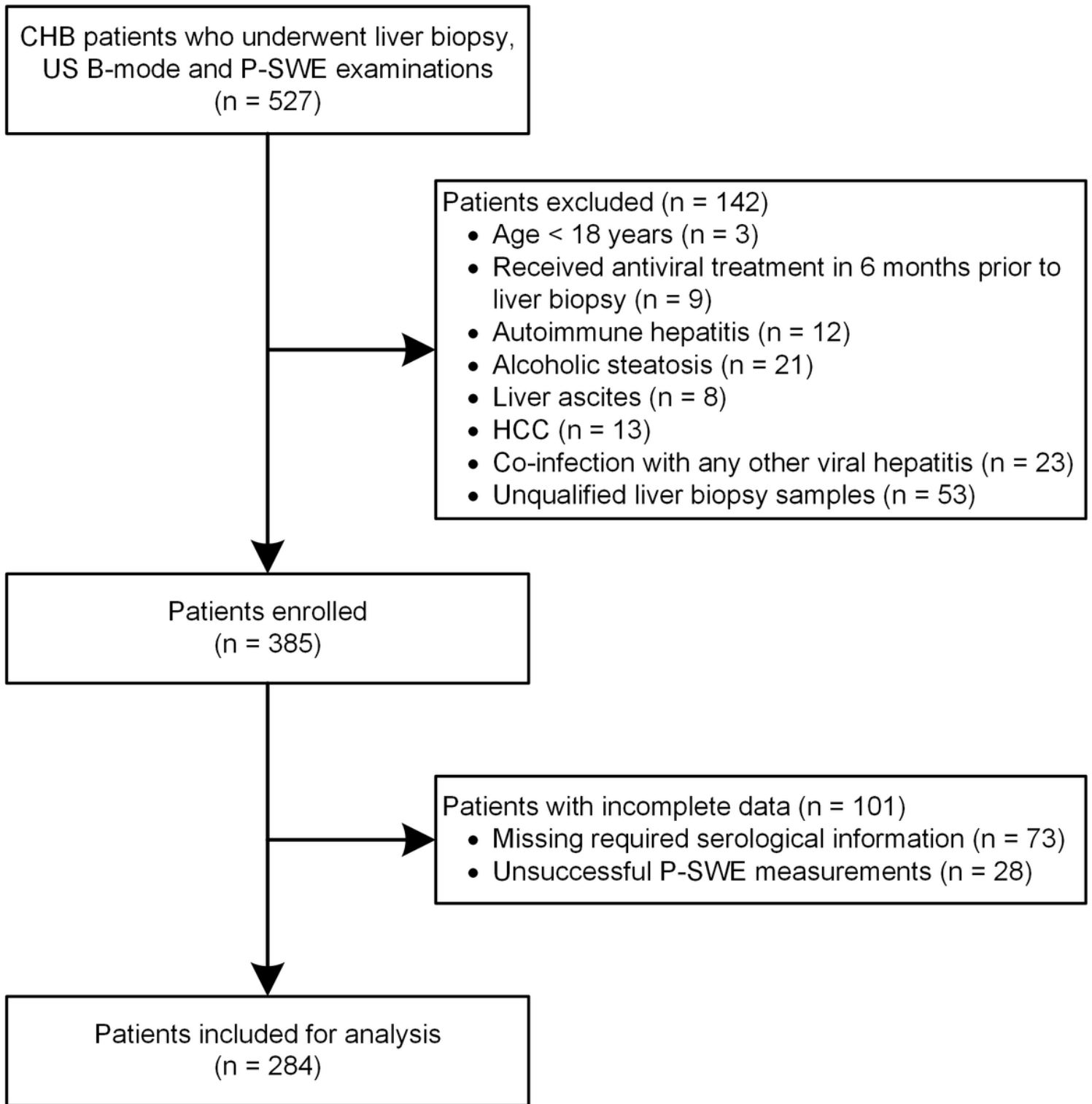


Figure 1

Workflow for patient screening.

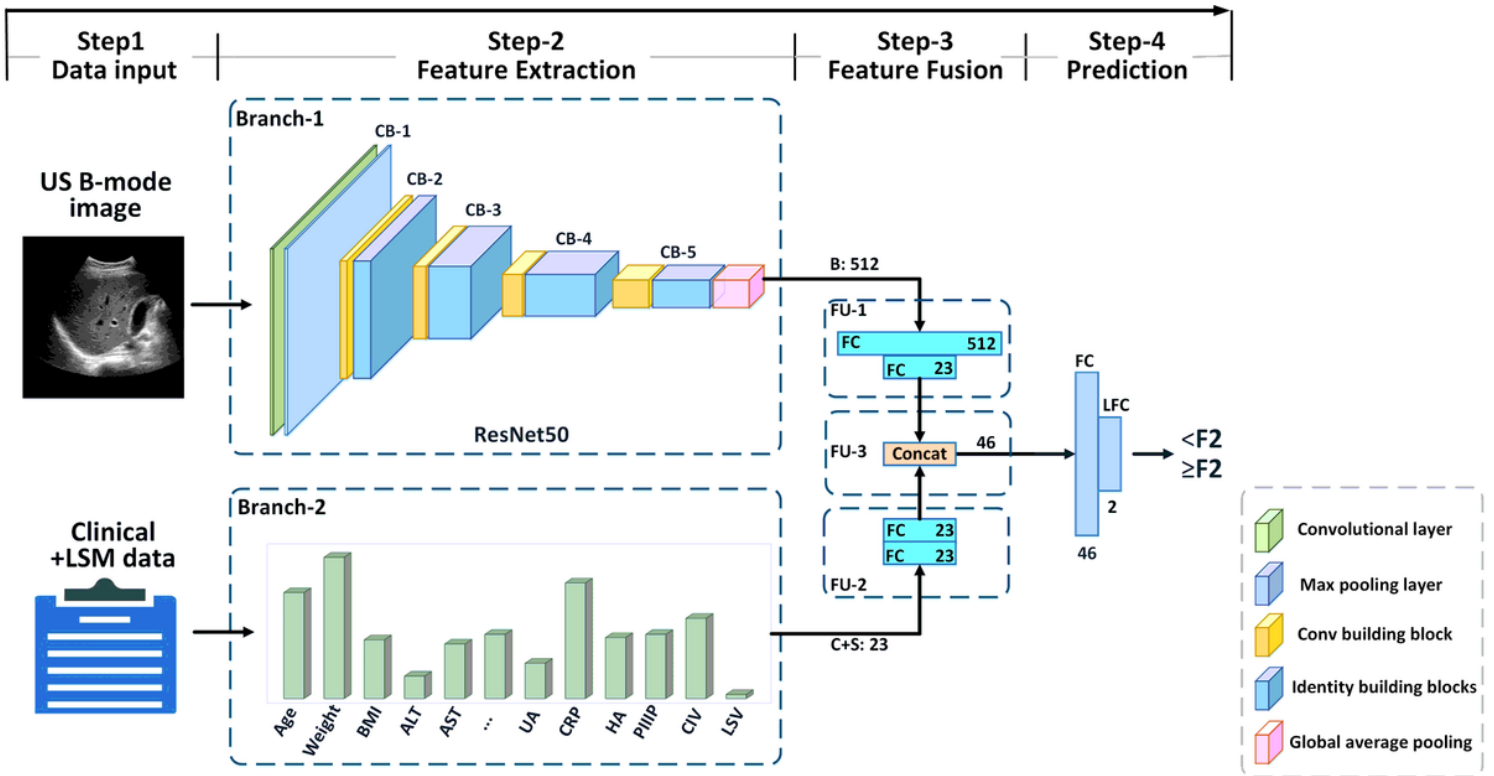


Figure 2

Architecture of DI-Net. US, ultrasound; LSM, liver stiffness measurement; FU, fusion unit; FC, fully connected; Concat, concatenation; LFC, last fully connected.

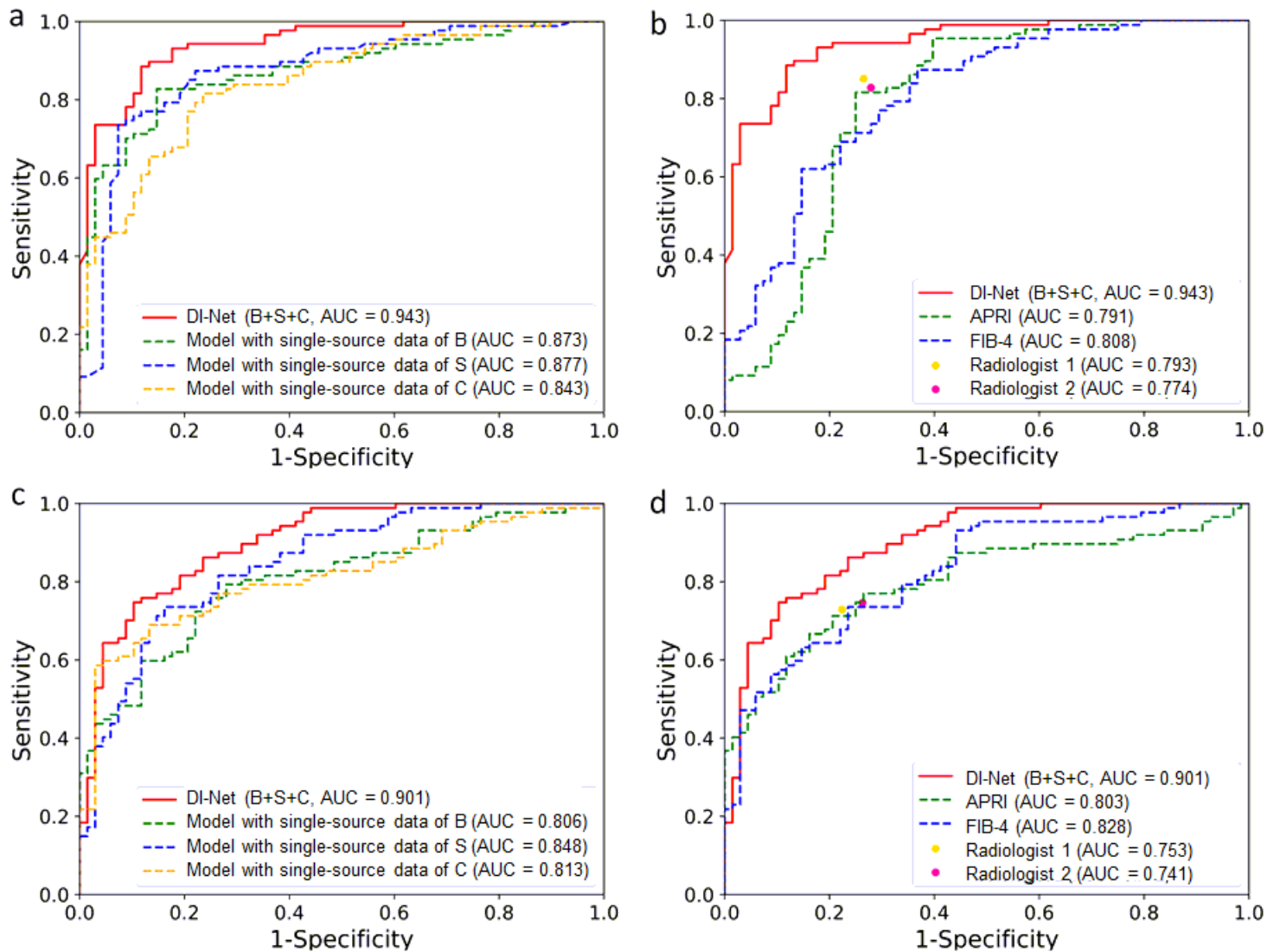


Figure 3

ROC curves yielded by different models. (a) and (b) are the cross-validation results, (c) and (d) are the external validation results. Note, for better comparison, only the cut-off points, instead of the entire ROC curves, were plotted for the radiologists' results. ROC, receiver operating characteristic; AUC, area under the ROC curve.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphicalAbstract.jpg](#)
- [Table2.jpg](#)
- [Table3.jpg](#)
- [Table1.jpg](#)

- [SupplementaryHEPI.docx](#)